特開2003-12250 リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

O

(19) 日本国格許庁 (J.P)

3 公费 點 华 噩 (I2)

蛤園2003-122509 (11)特許出願公開番号

. 35

	5333			
5 B 0 8	5 3 1 D	12/00		
5 B 0 6	304F	3/06	G06F	
デーマコート。(参う	11-		FI	
5H (2003, 4	平成15年4月25日(2003.4	(43)公開日		
2509A)	(P2003-122509A)	•		
166000	TAPATOR TOTAL			

認問中 304 533 531

> 3/06 12/00

G06F (51) Int.Cl.7

10 0

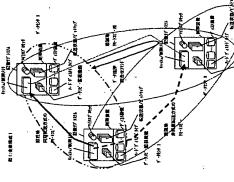
(条34月) 審査議決 未確決 指求項の数34 O.L.

(21)出順番母	特取2002~19971(P2002~19971)	(71) 出國人 000005108	000005108
(22)出版日	平成14年1月29日(2002.1.29)		株式会社日立製作所 東京都千代田区神田駿河台四丁目6番地
(31)優先権主張番号	(31)優先権主盟番号 特顧2001-240072(P2001-240072)	(72) 発明者	中野 俊夫 神奈川県小田原市中里322番地2号 株式
(32)優先日 (33)優先相主張国	平成13年8月8日(2001.8.8) 日本 (JP)	(72)発明者	会社日立製作所RAIDシステム事業部内中村 脚窓
			神奈川県小田原市中里322番地2号 株式会社日立製作所RAIDシステム事業部内
		(74) 代理人	100071283 弁理士 一色 既補
			· · · · · · · · · · · · · · · · · · ·

リモートコピー傾倒方法、これを用いた配像サブシステム、及び、これらを用いた広域データス トレージシステム (54) [発政の名称]

【瞑題】 3 以上のデータセンタ間で、常時、データ更新 の順序性を保持する。 (57) [政約]

これらのうち 1 つのデータセンタと、遠隔地に存在する 【解決手段】 付近地に存在する2つのデータセンタ間 第3のデータセンタとの間は、非同期リモートコピー機 能で連結し、付近地に存在する記憶サブシステムがホス トから受倒したデータの順序性を常時、保証しつつ第3 のデータセンタが、そのデータを保持する。更に、正常 **週月の際には直接にデータ転送を行なわない2つのデー** タセンタに散回された記憶サブシステムの間で、データ **伝送・受領・更新の進捗状態を把握する機能を、各記憶** は同期転送によるコピー機能を用いた接続構成とする。 サブシステムに設ける。



特開2003-12250 リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

σ

送するN個の転送先と、前記データを非同期転送するM 9を一時的に格納するキャッシュメモリと、これらを飼 **御するマイクロプロセッサにより、前紀データを同期転** 間の転送先とを有する記憶サブシステムのリモートコピ 【酌求項1】 制御情報を格納する制御メモリと、デー

前配制御メモリが、N+M個のデータの転送先に対応す る転送状態/ピットマップを格納する第1のステップ 前記転送状態/ピットマップのうち 1つに対応する転送 状態/ピットマップを有する別の記憶サブシステムであ のステップと、第2のステップの応答を受けて、転送状 1月 アントマップを更新する第3のステップとを有する データ更新の進捗状態を問合せるコマンドを発する第2 って、直接にデータ転送を行っていないものに対して、 ことを特徴とするリモートコピー制御方法。

第3のステップにおける転送状態/ピットマップの更新 む、データブロックの一部に対するデータの更新回数を は、前記データ更新の進捗状態を問合せるコマンドが含 **計数するためのカウンタ値を更新することを含むリモー** トコピー制御方法。

タを一時的に格納するキャッシュメモリと、これらを飼 **卸するマイクロプロセッサにより、前記データを同期転 改する転送先と、前記データを非同期転送する転送先と** 制御情報を格納する制御メモリと、デー を有する記憶サブシステムにおいて、

直接データの転送を行わない別の記憶サブシステムに対 し、データ更新の進捗状態を問合せる機能を有すること

を特徴とする記憶サブシステム。

自己と、直接データの転送を行わない別の配位サブシス を有し、前記進捗状態を問合せる機能により、当核1組 テムにおいて保持する、1組の転送状態/ビットマップ の転送状態/ピットマップを更新する記憶サブシステ 前配1組の転送状態/ピットマップは、データプロック

の一部に対するデータの更新回数を計数するためのカウ

ンタ値を格納する領域を有する記憶サブシステム。

第1のデータセンタに設配された第1の 2の記憶サブシステム、及び、第3のデータセンタに設 記憶サブシステム、第2のデータセンタに設置された第 **囮された第3の配憶サブシステムを有する広域データス**

第1の記憶サブシステムと第2の記憶サブシステムは、

非同期転送によりデータが転送され、第3の記憶サブシ ステムにおいて、当核転送されたデータの順序性が、常 時、保たれることを特徴とする広域データストレージン 第1の記憶サプシステムと第3の記憶サプシステムは、

節求項6配載の広域データストレージン [(()) |

第2の配憶サブシステムから、同期転送によりデータの 転送を受けた第1の記憶サブシステムが、第3の記憶サ ブシステムに対し、非同期転送により当該データを転送 する広域データストレージシステム。 ステムにおいて,

2

ステムにおいて、

第1の記憶サブシステムは、ホストから受倒したデータ を、同期転送により第2の配憶サブシステムへ転送する と共に、非同期転送により第3の配倣サブシステムにも 伝送する広域データストレージシステム

[節求項9] 顔求項6記載の広域データストレージシ ステムにおいて,

20

第1のデータセンタと第2のデータセンタは附近地に存 在し、第1のデータセンタと第3のデータセンタは遺隔 地に存在する広域データストレージシステム。 【酢氷項10】 第1のデータセンタに設配された第1 第2の記憶サブシステム、及び、第3のデータセンタに 設置された第3の記憶サブシステムを有する広域データ ストレージシステムにおける、3つ以上のデータセンタ の配位サブシステム、第2のデータセンタに設置された の間のリモートコピー鮭御方法であって、 25

ホストからのデータを第1の記憶サブシステムが受倒す る第1のステップと、 200

を、第2の記憶サブシステムへ、同期転送する第2のス 第1の配像サブシステムが、前配ホストからのデータ テップと、

を、第3の記憶サプシステムへ、非同期転送する第3の 第1の記憶サブシステムが、前配ホストからのデータ 35

へ到着したか否かを問合せる第4のステップとを有する く、 値配ホストからのデータが筑3の配像サブシステム 第2の記憶サブシステムから第3の記憶サブシステム 40

[哲米項11] 額米項10記載のリモートコピー飯御 方法において、更に、

リモートコピー制御方法。

転送する第6のステップとを有するリモートコピー制御 へ、第2の配像サブシステムが保持するデータの一部を 第1のデータセンタが機能停止する第5のステップと、 類2の記憶サブシステムから第3の記憶サブシステム 45

50 の記憶サブシステム、第2のデータセンタに設置された 【酢求項12】 第1のデータセンタに設置された第1

第2の記憶サブシステム、及び、第3のデータセンタに 段配された第3の記憶サブシステムを有する広域データ ストレージシステムにおける、3つ以上のデータセンタ の間のリモートコピー制御方法であって、

ホストからのデータを第1の記憶サブシステムが受領す 5第1のステップと、

05

を、第2の記憶サブシステムへ、同期転送する第2のス 第1の記憶サブシステムが、前記ホストからのデータ

第2の記憶サブシステムが、同期転送された前記ホスト からのデータを、第3の記憶サブシステムへ、非同期転 送する第3のステップと、

へ到着したか否かを問合せる第4のステップとを有する へ、前記ホストからのデータが第3の記憶サブシステム 第1の記憶サブシステムから第3の記憶サブシステム リモートコピー制御方法。

2

[韓求項13] 耐求項12記載のリモートコピー制御 方法において、更に、 第2のデータセンタが機能停止 する第5のステップと、

へ、第1の記憶サブシステムが保持するデータの一部を 第1の記憶サブシステムから第3の記憶サブシステム

伝送する第6のステップとを有するリモートコピー制御

を備え前記記憶資源に記憶されているデータが転送され 記憶資源に対するデータ書き込み手段 る複数の転送先が接続する配憶サプシステムにおけるリ モートコピー制御方法において、

記憶サブシステムが、前記記憶資源にデータを審き込む

記憶サブシステムが、前記記像資源における前記データ と、データの뿁き込み順に付与されるシーケンス番号と 啓き込みが行われた位置を特定する書き込み位置情報 の対応づけとを記憶するステップと、

記憶サブシステムが、前記データ售き込みにより書き込 んだデータと前配対応づけとを前記転送先に送信するス テップと、 記憶サブシステムが、前記転送先から送られてくる前記 記憶サブシステムが、記憶している前記対応づけと前記 前配転送先において未反映となっている曹き込みデータ 転送先から受信した前記シーケンス番号とに基づいて、 シーケンス番号を受信するステップと、

を有することを特徴とするリモートコピー制御方法。 を把握するステップと、

20 のサイトに設置された第3の記憶サブシステムと、を有 【間求項15】 第1の記憶資源に対するデータ魯き込 み手段を備え第1.のサイトに設置された第1の記憶サブ システムと、第1の記憶サブシステムに接続するホスト 第3の記憶資源に対するデータ書き込み手段を備え第3 と、第2の記憶資源に対するデータ書き込み手段を備え 第2のサイトに設置された第2の記憶サブシステムと、

する広域データストレージシステムにおけるリモートコ ピー制御方法において、 **第1の記憶サブシステムが、前記ホストからの指示によ** り第1の記憶資源に対してデータの書き込みを行うステ 第1の記憶サブシステムが、第1の記憶資源における前 記データ番き込みが行われた位置を特定する書き込み位 **置情報と、データの書き込み順に付与されるシーケンス** 番号との対応づけとを記憶するステップと、 第1の記憶サプシステムが、前記データ書き込みにより **費き込んだデータと前記対応づけとを第2の記憶サブシ** ステムに送信するステップと、 0

とを受信して前記データを第2の記憶資源に記憶し、前 記データと前記対応づけとを第3の記憶サブシステムに 第2の記憶サブシステムが、前記データと前記対応づけ

第3.の記憶サブシステムが、前記データと前記対応づけ とを受信して前記データを第3の記憶資源に記憶すると ともに前記対応づけにおける前記シーケンス番号を第1

第1の記憶サプシステムが、前記シーケンス番号を受信 して、前記シーケンス番号と記憶している前記対応づけ の記憶サプシステムに送信するステップと、 20

とに基づいて、第3の記憶資源に未反映となっているデ ータを把握するステップと、

【醋求項16】 前記第2の記憶サブシステムが降害等 を有することを特徴とするリモートコピー制御方法。 25

前記第1の記憶サブシステムが、前記受信した前記シー こより使用できなくなった場合に、

ケンス番号と前記対応づけとに基づいて把握した前記第 3の記憶資源において未反映となっている差分のデータ とその書き込み位置情報とを前記第3の記憶サブシステ ムに送信するステップと、 30

書き込み位置情報とを受信して、前記差分データを前記 前記第3の記憶サブシステムが、前記差分データと前記 書き込み位置情報に基づいて前記第3の記憶資源に記憶 して前記第1の記憶資源と前記第3の記憶資源の内容を 同期させるステップと、 35

り前記第1の記憶資源に対してデータ番き込みが行われ タと前記第1の記憶資源における前記データ書き込みが 行われた位置を特定する書き込み位置情報とを前記第3 第1の記憶サブシステムが、前記ホストからの指示によ た場合に、前記データ書き込みにより書き込まれたデ の記憶サプシステムに送信するステップと、 40

前記第3の記憶サブシステムが、前記データと前記書き 込み位置情報とを受信して前記データを前記第3の記憶 資源の前記書き込み位置情報で特定される位置に記憶す るステップと、

を有することを特徴とする醋求項15に記載のリモート コピー制御方法。 【醋求項17】 前記第2の記憶サブシステムが回復し

前記第1の記憶サブシステムが、前記第1の記憶資源に 答納されている全データを前記第2の記憶資源に転送す **耳び使用できるようになった場合に、**

第1の記憶サブシステムが、前記ホストからの指示によ データとシーケンス番号とを第2の記憶サブシステムお 第2の記憶サブシステムが、前記データと前記シーケン り第1の記憶資源にデータ書き込みを行い、書き込んだ よび第3の記憶サブシステムに送信するステップと、

ス番号とを受信して前記データを第2の記憶資源に記憶 対するデータ書き込みが行われた位置を特定する昔き込 前記第2の記憶サブシステムが、前記第2の記憶資源に み位置情報と、データの書き込み頃に付与されるシーケ

ンス番号との対応づけを記憶するステップと(位置情報

前記第3の記憶サブシステムが、記憶サブシステム1か ら送られてくる前記データと前記シーケンス番号とを受 もに前記対応づけにおける前記シーケンス番号を前記第 冒して前紀データを前紀第3の記憶資源に記憶するとと

2の記憶サブシステムに送信するステップと、

前記第1の記憶サブシステムから前記第3の記憶サブシ 前記第2の記憶サブシステムが、記憶サブシステム3か ら送られてくる前記シーケンス番号を受信して、前記シ ーケンス番号と自身が記憶している前記シーケンス番号 とこれに対応する魯き込み位置情報とに基づいて、前記 ステムへのデータの転送を停止するステップと、

前記第2の記憶サブシステムが、前記把握した前記第3 の記憶資源において未反映となっている差分のデータと その書き込み位置情報とを前記第3の記憶サブシステム に送信するステップと、

第3の記憶資源に未反映となっているデータを把握する

前記第3の記憶サプシステムが、前記差分データと前記 書き込み位置情報とを受信して、前記差分データを前記 勘き込み位置情報に基づいて前配第3の記憶資源に記憶 して前記第1の記憶資源と前記第3の記憶資源の内容を 同期させるステップと、

を有することを特徴とする間求項16に記載のリモート コピー制御方法。

【開求項18】 記憶資源に対するデータ售き込み手段 を備え前記記盤資源に記憶されているデータが転送され る複数の転送先が接続する記憶サブシステムにおけるリ モートコピー制御方法において、

記憶サブシステムが、前記記憶資源にデータを書き込む

記憶サプシステムが、前記記像資源における前記データ と、データの書き込み順に付与されるシーケンス番号と 音き込みが行われた位置を特定する書き込み位置情報 の対応づけとを生成するステップと、

記憶サブシステムが、前記データ書き込みにより書き込 んだデータと前記対応づけとを前記転送先の一つである 転送先Aに送信するステップと

転送先Aが記憶サプシステムから送信されてくる前記対 応づけを受信して記憶するステップと、 転送先Aが別の前記転送先である転送先Bから送られて

伝送先Aが、記憶している前記対応づけと転送先Bから 受信した前記シーケンス番号とに基づいて、転送先Bに おいて未反映となっている書き込みデータを把握するス くる前記シーケンス番号を受信するステップと、

【請求項19】 第1の記憶資源に対するデータ書き込 を有することを特徴とするリモートコピー制御方法。

み手段を備え第1のサイトに設配された第1の記憶サブ 第3の記憶資源に対するデータ書き込み手段を備え第3 システムと、第2の記憶資源に対するデータ書き込み手 段を備え第2のサイトに設置された第2の記憶サブシス のサイトに設置された第3の記憶サブシステムと、を行 する広域データストレージシステムにおけるリモートコ テムと、第2の記憶サプシステムに接続するホストと、 12 20

り第2の記憶資源に対してデータの皆き込みを行うステ 第2の記憶サプシステムが、前記ホストからの指示によ ピー制御方法であって、

情報とデータの書き込み順に付与されるシーケンス番号 前記書き込みが行われた位置とを特定する書き込み位置 第2の記憶サブシステムが、前記書き込んだデータと、 とを第1の記憶サブシステムに送信するステップと、 25

第1の記憶サブシステムが前記データと前配シーケンス。 番号とを受信してこれらを第1の記憶資源に記憶し、前 記シーケンス番号と、前紀データが書き込まれた第1の 記憶資源上の格納位置を特定する書き込み位置情報とを 対応づけて記憶するステップと、 30

第2の記憶サブシステムが、前記データ費き込みにより 番き込んだデータと前記シーケンス番号とを第3の記憶 サブシステムに送信するステップと、 32

るとともに前記データと対になっていた前記シーケンス 第3の記憶サブシステムが前記データと前記シーケンス 番号とを受信して前記データを第3の記憶資源に記憶す

第1の記憶サブシステムが、前記シーケンス番号を受信 して、前記シーケンス番号と記憶している前記対応づけ とに基づいて、第3の記憶資源に未反映となっているデ 番号を第1の記憶サブシステムに送信するステップと、

【静求項20】 前記第2の記憶サブシステムが障害等 を有することを特徴とするリモートコピー制御方法。 ータを把握するステップと、 45

前記第1の記憶サブシステムが、前記受信した前記シー により使用できなくなった場合に、

ケンス番号と前記対応づけとに基づいて把握した前記第 3の記憶資源において未反映となっている差分のデータ 20

とその母き込み位置情報とを前記第3の記憶サブシステ

前配第3の配像サブシステムが、前配差分データと前配 敬き込み位置情報に基づいて前配第3の記憶資源に記憶 して前記第1の記憶資源と前記第3の記憶資源の内容を むき込み位配情報とを受信して、前記差分データを前記 同期させるステップと、

第1の記憶サブシステムが、前記ホストからの指示によ 行われた位置を特定する事き込み位置情報とを前配第3 り前配第1の配像資源に対してデータ母き込みが行われ た場合に、前紀データ敬き込みにより敬き込まれたデー タと前記第1の記憶資源における前記データ母き込みが の配憶サブシステムに送信するステップと、

込み位置情報とを受信して前配データを前配第3の記憶 前記第3の記憶サブシステムが、前記データと前記告き 資源の前記母き込み位置情報で特定される位配に記憶す るステップと、を有することを特徴とする間求項19に 記載のリモートコピー制御方法。

【朝求項21】 前記第2の記憶サブシステムが回復し 再び使用できるようになった場合に、

前配第1の記憶サブシステムが、前記ホストにより傷き 込まれた前記第1の記憶資源に格納されている全てのデ 一夕を前記第2の記憶資源に転送するステップと、

第1の記憶サブシステムが、前記ホストからの指示によ り第1の記憶資源に対してデータ母き込みを行うステッ

費き込んだデータと前配シーケンス番号とを第2の記憶 第1の記憶サプシステムが、前記データ巷き込みにより サブシステムおよび第3の記憶サブシステムに送信する 第2の記憶サブシステムが、前記デー久と前配シーケン ス番号とを受信して前記データを第2の記憶資源に適時 に記憶するステップと、

対するデータ街き込みが行われた位置を特定する街き込 前配第2の配位サプシステムが、前記第2の配位資源に み位置情報と、データの뿁き込み順に付与されるシーケ ンス帝号との対応づけを記憶するステップと、

ンス番号を前配第2の記憶サブシステムに送倡するステ 顔に配像するとともに前配対応づけにおける前配シーケ 前配第3の記憶サブシステムが、前配データと前配シー ケンス番号とを受信して前記データを前記第3の記憶資

前配第2の記憶サプシステムが、前配シーケンス番号を **受信して、前記シーケンス番号と記憶している前記対応 信記第1の記憶サブシステムから前記第2の記憶サブシ** づけとに基づいて、前配第3の配憶資源に未反映となっ ステムへのデータの転送を停止するステップと、

ケンス番号と前配対応づけとに基づいて把握した前記第 前配第2の記憶サブシステムが、前配受信した前配シー ているデータを把握するステップと、

3の記憶資源において未反映となっている差分のデータ

ひき込み位置情報とを受信して、前記差分データを前記 前記第3の記憶サブシステムが、前記差分データと前記 你き込み位配情報に基づいて前配第3の配憶資源に記憶 して前記第1の記憶資源と前記第3の記憶資源の内容を 同期させるステップと、 05

コピー制御方法。

【耐求項22】 前記第1の記憶サブシステムと前記第 2の記憶サプシステムとの間は同期転送により運用さ に記載のリモートコピー制御方法。 2

【櫛求項23】 前記第2の記憶サブシステムもしくは の記憶資源もしくは前記第3の記憶資源における格納位 前記第3の記憶サブシステムは、前記第1の記憶サブシ ステムにおける哲き込み位置竹報に対応づけた前配第2 啞にデータを記憶することを特徴とする、即求項15∼ 17および19~21のいずれかに記載のリモートコピ 一制御方法。

25

30

システムBの中から記憶サブシステムAの処理を代行さ せる記憶サプシステムbを選出するステップと、

35

を備えることを特徴とするリモートコピー制御方法。 40

【
即求項26】 前記障害が発生したことを検知するス 記憶サブシステムトを選出する前記ステップを、 求項24に記載のリモートコピー制御方法。 45

とその醬き込み位置情報とを前記第3の記憶サブシステ

ムに送信するステップと、

を有することを特徴とする臨求項20に配載のリモート

システムとの間は非同期転送により運用されることを特 れ、前記第1の記憶サブシステムと前記第3の記憶サブ **徴とする前求項15~17および19~21のいずれか**

ホストが接続された配像サブシステム Aとこれに通信手段を介して接続する複数の他の記憶サ プシステムBと、を有する広域データストレージシステ ムにおけるリモートコピー制御方法であって、 [耐求項24]

るステップであることを特徴とする前求項24に記載の

リモートコピー制御方法。

[間求項30] 配位資源に対するデータ番き込み手段 を備え前記記憶資源に記憶されているデータが転送され る複数の転送先が接続する配位サプシステムにおける配

サブシステムBが記憶しているデータの内容を一致させ

り、前記選出した記憶サブシステムbとこれ以外の記憶

テムBに記憶しているデータの差分データの複写によ

記憶サブシステムBが、前記ホスト、記憶サブシステム A、前記通信手段のうち少なくともいずれか一つに障害 が発生したことを検知するステップと、

記憶サブシステムBが、障査を検知した場合に記憶サブ

記憶サブシステムBが、選出した記憶サブシステムbと これ以外の記憶サブシステムBが記憶しているデータの 配憶サプシステムbに接続する剧ホストにより前記ホス 内容を一致させるステップと、記憶サブシステムBが、 トの運用を引き継ぐステップと、

テップ、および記憶サプシステムAの処理を代行させる 上の前記記憶サプシステムBが行うことを特徴とする前

対して送信されるハートビートメッセージがあらかじめ **前配記像サブシステムAから前記記像サブシステムBに** 散定された時刻に受信できない場合に障害が発生したと

する広域データストレージシステムにおける耐水項15 **に記載のリモートコピー制御方法に使用する値配第1の** 記憶サプシステムとして機能する記憶サプシステムであ 前記ホストからの指示により第1の記憶資源に対してデ **一夕の街き込みを行う手段と、**

前記各配億サプシステムBに記憶しているデータの更新

節サブシステムbを選出する前記ステップが、

順序を示すシーケンス番号のうち最新のシーケンス番号 同士を比較して、これらのうちで最新の前配シーケンス 番号を記憶している記憶サブシステムBを、前記記憶サ プシステムAの処理を代行させる前記記憶サプシステム

ムBの中から記憶サプシステムAの処理を代行させる記 bとして選出するステップであることを特徴とする 即求

【醇求項28】 前配各配億サブシステムBに配億して いるデータの更新順序を示す前記シーケンス番号に抜け がある場合に、連続しているシーケンス番号のうち段策

項24に記載のリモートコピー制御方法。

第1の配位資源における前記データ母き込みが行われた 位置を特定する事き込み位置情報と、データの費き込み 低に付与されるシーケンス番号との対応づけを配位する 9 前配データ母き込みにより母き込んだデータと前記対応 3の配憶サブシステムから送信されてくるシーケンス街 母を受信して、このシーケンス番号と記憶している前記 対応づけとに基づいて、第3の配物資源に求反映となっ づけとを第2の記憶サブシステムに送信する手段と、第 2

のシーケンス番号を用いて前記比較を行うことを特徴と

外の記憶サブシステムBが記憶しているデータの内容を

前記記憶サプシステムbとこれ以外の記憶サプシステム との間で、全データの複写、もしくは、各記憶サブシス

一致させる前記ステップが、

を備えることを特徴とする配位サブシステム。 ているデータを把握する手段と、

る複数の転送先が接載する配像サブシステムAにおける 【間求項32】 配位資源に対するデータ仍き込み手段 を備え前記記憶資源に記憶されているデータが転送され 酢求項18に配載のリモートコピー制御方法に使用する 記憶サブシステムAから送られてくる、前記記憶資源に 対して告き込まれたデータと、そのデータの告き込みが 行われた位置を特定する语き込み位置情報とデータの費 き込み順に付与されるシーケンス番号との対応づけとを 前配転送先として機能する配憶サブシステムであって、 受信して記憶する手段と 20 25

別の前配転送先である転送先Bから送られてくる前配シ ーケンス番号を受信する手段と、

記憶している前記対応づけと転送先Bから受信した前記 シーケンス番号とに基づいて、転送先Bにおいて未反映 となっている都き込みデータを把握する手段と、 30

求項14に記載のリモートコピー制御方法に使用する前

配配像サプシステムであって、

に付与されるシーケンス番号との対応づけとを記憶する

前記データ費き込みにより費き込んだデータと前配対応 づけとを前記転送先に送信する手段と、前記転送先から

前記記憶資源における前記データ母き込みが行われた位 団を特定する鸖き込み位置情報と、データの勘き込み順

前配配位資源に対してデータ都き込みを行う手段と、

配像している前記対応づけと前配板送先から受信した前

送られてくる前記シーケンス番号を受信する手段と、

記シーケンス番号とに基づいて、前記転送先において未

反映となっている仍き込みデータを把握する手段と、

を備えることを特徴とする記憶サブシステム。

み手段を備え第1のサイトに設置された第1の記憶サブ システムと、第1の記憶サブシステムに接続するホスト

を備えることを特徴とする記憶サブシステム。

【朝求項33】 第1の記憶資源に対するデータ書き込 み手段を備え第1のサイトに数配された第1の記憶サブ システムと、第2の記憶資敵に対するデータ曲き込み手 段を備え第2のサイトに設置された第2の記憶サブシス のサイトに設置された第3の記憶サブシステムと、を有 する広域データストレージシステムにおける加泉項19 第3の配位資源に対するデータ母き込み手段を仰え第3 **に記載のリモートコピー館御方法に用いられる値配等1** テムと、第2の記憶サブシステムに接続するホストと、 9 35

トからの指示により敬き込んだデータと、前記歌き込み 類2の記憶サブシステムから送信されてくる、値配ホス が行われた位配とを特定する母き込み位配桁組とデータ の敬き込み順に付与されるシーケンス都号との対応づけ とを受信して、これを第1の配位資源に記憶する手段

の配位サプシステムとして機能する配位サブシステムで

20

第3の記憶資源に対するデータ番き込み手段を備え第3 のサイトに設置された第3の記憶サブシステムと、を有

と、第2の記憶資源に対するデータ母き込み手段を備え 第2のサイトに設置された第2の記憶サブシステムと、

認識するステップであることを特徴とする間求項24に

20

第3の記憶サブシステムから送られてくるシーケンス番 いる前記対応づけとに基づいて、第3の記憶資源に未反 号を受信して、このシーケンス番号と、自身が記憶して を備えることを特徴とする記憶サブシステム。 映となっているデータを把握する手段と、

【蘭求項34】 ホストが接続された記憶サブシステム Aとこれに通信手段を介して接続する複数の他の記憶サ プシステムBと、を有する広域データストレージシステ ムにおける酢水項24に記載のリモートコピー制御方法 に使用する前記記憶サブシステムBとして機能する記憶 サブシステムであって

前記ホスト、記憶サプシステムA、前記通信手段のうち

少なくともいずれかー

障害を検知した場合に記憶サブシステムBの中から記憶 サブシステムAの処理を代行させる記憶サブシステムb つに障害が発生したことを検知する手段と、 を選出する手段と、

選出した記憶サブシステムもとこれ以外の記憶サブシス テムBが記憶しているデータの内容を一致させる手段

記憶サブシステムbに接続する副ホストにより前記ホス

を備えることを特徴とする記憶サブシステム。 トの運用を引き継ぐ手段と

[発明の詳細な説明]

[0001]

볎装置の障害が生じた後に、速やかに、その障害から復 置され、相補的な動作を行う3つ以上の外部記憶装置か 【発明の属する技術分野】本発明は、災害による外部記 外部記憶装置が相互に100kmから数百km隔てて設 旧可能な広域データストレージシステムに係り、特に、 らなる広域データストレージシステムに関する。 [0002]

は非同期にて行うことが開示されている。また、本件の 【従来の技術】本件の出願人による特開平11-338 647号公報には、システムとデータの2重化を同期又 非同期で遠隔地にデータのコピーを行う技術が開示され 出願人による特開2000-305856号公報には、

[0003] このように、本件の出願人は、大型計算機 システム、サーバー、ネットワーク上のパーソナルコン ピュータ、その他の上位計算機システム(以下、ホスト ータを、遠隔地に設置された第2の記憶サブシステムに 対し、そのデータの順序性を常時保証しながら、非同期 という。)から、データの順序を特定する特別な制御情 報を受領することなく、そのデータを受け取った外部記 **施装置(以下、記憶サブシステムという。)が、そのデ 告き込むという、非同期リモートコピーの技術を所有し** 伝送により第2の記憶サブシステムへ連続して間断なく

20 【0004】また、同期転送の技術を用いてコピーを行

近地又は遠隔地に設置された記憶サブシステムとの間の コピー制御が連動するため、巨視的にみて常に2つの記 うときは、ホストとこれに接続された記憶サプシステム との間のデータ更新処理と、この記憶サブシステムと付 僚サブシステム間でデータが一致しており、その書込み 順序性も同時に保証されている。尚、適当なデータ転送 経路を選択すれば、2つの記憶サブシステムの距離が1 00kmを超える場合であっても、同期転送によるコピ 09

【0005】昨今、データを安全に格納し保持すること が重要であるという認識が高まっており、データストレ ージの市場では、ディザスタリカバリシステムを要請す る声が多く聞かれる。従来のように、データ格納拠点を 2つ設け、かかる2地点間を同期転送又は非同期転送で 結ぶことは実現されている。しかし市場は、第3、第4 のデータ格納拠点(以下、データセンタという。)を要 求し、これらの間での完全な又は完全に近いディザスタ リカバリシステムの構築を望んでいる。 2 2

【0006】その理由は、3拠点以上のデータセンタを に、残る複数のセンタ間でデータ冗長化の回復・維持が 設置しておけば、これらのうち 1個所が災害に見舞われ ても、引き続き発生する災害のリスクを軽減するため 図れるであろうという期待にある。 [0007] 従来の技術では、3以上のデータセンタを **構築した場合に、ホストから受領する1 / ○を唯一の記 憶サブシステムの論理ポリュームで受領し、これを複数** のデータセンタヘリモートコピー技術を用いて転送する 際の配慮が十分で無かった。例えば、一つのデータセン タが災害によりダウンした場合に、残る2以上のデータ センタ間で、データの順序性を保証した論理ポリューム を構築できるか、更新状態を引き継ぎデータの不整合を ーを可能とするシステムの再構築ができるかといった問 無くすことができるか、附近地又は遠隔地に対するコピ 題に関し、配慮が足りなかった。 22 30

【発明が解決しようとする課題】災害はいつ発生するか 不明なため、3以上のデータセンタ間で、常時、データ 更新の順序性を保持しなければならない。 [8000] 35

遼サブシステムへ配信し、かつ、如何なる時点で災害が タセンタの記憶サブシステムで、常時、保証する広域デ ュームが受領したデータを、遠隔地又は附近地の別の記 ず、複数のリモートコピー構成を連結し、同一論理ボリ [0009] このため、ホストに特殊な機能を具備せ 発生しても、ホストからのデータの更新順序を、各テ ータストレージシステムを構成しなければならない。 40

[0010] 本発明に係る広域データストレージシステ ムでは、記憶サブシステムの内部に、冗長化した論理ポ リュームを設けることなく、別の記憶サブシステムに対 **ン、データをコピーすることにより上記の製題を解決し**

記憶サブシステムが保持しているデータの差分のみを転 作業(再同期、リシンク)において、災害発生直前に各 【0011】また、本発明に係る広域データストレージ で、管理情報を置り取りし、データの更新状態を各記憶 サブシステムで監視し管理する。そして、災害後の復旧 送することで、即時に、ホストの交代(failove ンステムでは、災害後の復旧作業として、広域ストレー ジシステムの再構成を想定しており、正常な運用時に、 直接、データ転送を行っていない記憶サブシステム間

[0012] <データ更新の順序性を常時保証すること について>ここで、順序性の保持の時間的範囲について r)と、アプリケーション実行の継続を行う。 桶足説明する。

[0013] ホストから発行された1/0は記憶サブシ ステムに書き込まれ、記憶サブシステムが報告するデー タの沓き込み完了報告を認識し、ホストは次のステップ を実行する。ホストは記憶サブシステムのデータ掛き込 こみの順序性は、ホストが記憶サブシステムから書き込 次の1/〇を正常には発行しない。従ってデータの皆き み完了報告を受領する前後で、記憶サブシステムが順序 性保存の何らかの処理をすることで維持されるべきもの み完了を受領しない場合又は障害報告があった場合は、

ピーされるデータが付近地又は遠隔地(以下、単に「別 理、及びデータ転送処理時間が長くかかり、性能が遅延 【0014】 同期転送のリモートコピーでは転送されコ 別地の記憶サプシステムからの書き込み完了を受領した 後、ホストに対し書き込み完了報告を行う。リモートコ ピーを行わない場合と比較し、リモートコピーに係る処 データ転送に伴う処理時間が増大し、リモートコピーを 行うことによりホストの I / O処理の性能をさらに低下 地」と略記する。)の記憶サプシステムに書き込まれ、 する。リモートコピーにおける接続距離を延長すると、 させる。これを打破する一つの方法が非同期転送であ

--タ転送を行ない別地の記憶サブシステムの書き込み完 [0015] 非同期転送は、ホストから1/0を受領し た配像サプシステムが、別地の記憶サブシステムへのデ 7を待たずに、ホストから1/0を受領した記憶サブシ り、記憶装置サブシステム間のデータ転送は、ホストの 1/0処理と関係が無くなり、ホストの1/0処理と非 込まなければ、別地の記憶サブシステムのデータ順序性 を来す可能性がある。データの順序性を常時保証する機 同期に実行できる。しかし、ホストからのデータの到着 順序を守って、別地の記憶サブシステムヘデータを書き は維持されず、両記憶サブシステム間でデータの不整合 ステムが書き込み完了報告をホストへ返す。これによ 能を追加すれば、このような可能性を極小化できる。

[0016] 別地の記憶サブシステムは、ホスト1/0 を受領した記憶サプシステムと比較し、通常はデータの

って記憶サブシステムへ書き込む限り、データの順序性 に矛盾は無く、ジャーナルファイルシステムやデータベ 更新は遅れているが、ホストからのデータ到溶瓶序を守

【0017】一方、データの順序性を維持せず、別地の 記憶サブシステムヘリモートコピーしてデータを反映さ せる方法もある。この方法は、ある時点までのホストか テムへ纏め書きする。ある時点までのデータ書き込みが 終わった段階で、データ転送を終了し、以降、次回の額 め書きまで、リモートコピーのデータ転送を抑止し、抑 ら受領したデータを別地へ送り、それらを記憶サブシス 止している間のデータ順序性、ホストから受領した1/ 一スリカバリ処理により、障害時の回復が可能である。 0の一貫性を保証する。

る機能が不要であるが、ある程度の更新分のデータを溶 込みが全て完了した段階で、データ整合性を保証してい る。この方法ではリモートコピーを行っている間に障害 【0018】この方法では、データの順序情報を付与す えておいて、その更新分を一括転送し、リモートへ書き が発生すると、リモート側のデータ更新は順序性を維持 して更新されていないため全域となる恐れがある。リモ **ートコピーのデーク転送を止めている間のみ、デーク整** 合性を保証でき、adapliveと呼ばれている。 2

データ順序情報をプロック毎に管理する措置を施してい。 れば、ホストに完了報告を返す際に、記憶サブシステム がデータの順序性を保証する処理をしていることが特徴 である。記憶サプシステムの制御装置におけるオーバへ ッドや内部処理の遅延時間に拘らず、ホストに返す際に 【0019】出願人の所有する"データの順序性を常時 **保証する非同期転送によるリモートコピー"の技術によ** 25

している。この一方で、リモート側の記憶サブシステム 【0020】実際には、ホストから受領する1/0発行 間隔よりかなり短い時間で、ブロック毎の制御・管理を は、出願人のリモートコピーの技術が、データに順序情 ローカルノリモートのデータ更新の時間差が、例え半日 あっても、順序性さえ正しければ、更新データ全てを映 報を付与してデータブロックを転送し、これに基づきデ でデータ配信を待ちきれずタイムアウト (Timeout) と **一夕の順序を守って書き込みを行なっている点である。** する値は、1時間以上に設定可能でもある。大切なの 失してしまう不整合より良いからである。 るため、常時順序性を保証できる。 30 35

に転送可能な転送経路、所定の管理情報の避り取りが可 【課題を解決するための手段】データを同期及び非同期 [0022] ここで、更新進捗管理手段は、各記憶サブ 能な通信線路、及び、デーク更新進捗管理手段により、 3 以上のデータセンタを相互に連結する構成とする。

システムにおけるデータ更新の進捗状態を、適宜、監視 システムに設けられ、いつ発生するか分からない災害に 対応するため、他のデータセンタに設置された記憶サブ 20

リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

相互にその記憶サブシステムのデータ更新状態を把

【0023】具体的には、直接データ転送を行なってい 掘させる手段である

ない配憶サブシステムの各々が、転送状態/ビットマッ タ更新 (リモートコピー) の進捗を監視し管理する機能 か、一方が問合せ、他方がこれに応答することで、デー プを持ち、転送プロックのどの位配が何回更新された

05

[0024]

2 【発明の実施の形態】3以上のデータセンタに、それぞ れ設置された記憶サプシステムの間を、同期転送により

て、1箇所のプライマリデータセンタの記憶サブシステ ムから、これを除いた残りの別拠点の2以上のデータセ ンタの各記憶サブシステムへ、ホストからプライマリの 【0025】かつ、データの順序性を常時、連続的に保 記憶サブシステムが受倒したデータを、ホストが更新し 証する非同期リモートコピーの技術で連結する。そし た順序を保持しつつ、連続的に転送し格納する。

[0026] データが、ホストからの更新順を保証して 冗長構成化されるため、万一、データセンタに災害・障 **事が発生しても、残ったデータセンタの配億サブシステ** ムの間で、各記憶サブシステム間の差分データのみを転 **復でき、又は、デーク喪失を最小限度とすることができ** 送することで、即時に、リモートコピーの運用構成を回

5、図6を用いて同期低送によるコピー又は非同期リモ [0027] <同期・非同期について>まず始めに、図 **ートコピーを定義する。**

してから、ホストに更新処理の完了を報告する処理手順 [0028] 同期転送によるコピーとは、ホスト1から 記憶サブシステム1に、データの更新(协込み)指示が有 った場合に、その指示対象が附近地に設置された記憶サ プシステム2にも仏込むデータであるときは、記憶サブ をいう。ここで、附近地とは、いわゆるメトロポリタン ネットワークと称される100km程度までの範囲を言 システム2に対して、指示された更新(街込み)が終了 うものとする。

これを記憶サブシステム1で受領し(圖)、最後に う(■)。途中の処理に失敗した場合には、ホスト1に 5)では、ホスト1から受領した更新データプロックを ホスト1に対し更新データブロックの働き込み完了を行 記憶サブシステム1で受領し(🔳)、そのデータブロッ クを記憶サプシステム2に転送し(■)、 むき込み完了 [0029] つまり、同期転送のリモートコピー(図 助き込み障害を報告する。

内容が、巨視的にみて常に一致している。このため、災 [0030] 同期転送によるコピーを実施すると、ホス ト1に接続された近い方の記憶サブシステム1と、附近 地に股団された違い方の記憶サブシステム2のデータの

る効果がある。尚、巨視的にみて常に一致とは、同期転 の単位 (μ sec, msec) で、一致していない状態が有り **むにより一方がその機能を失った場合であっても、他方** されているので、残るシステムで迅速に処理を再開でき 送の機能を実施中には、制御装置や電子回路の処理時間 得るが、データ更新処理完了の時点ではデータは必ず同 一の状態になっていることを意味している。これは、附 近地の記憶サブシステムへの更新データの反映が終了し ない限り、ホストに近い倒の記憶サプシステム1の更新 の記憶サブシステムに災害直前までの状態が完全に保存 処理を完了できないためである。

は、ホスト1からこれに接続された近い方の記憶サブシ [0031] 一方、非同期リモートコピー(図6)と

断処理が終わり次第、ホスト1に対し更新処理の完了を にも替込むデータであっても、記憶サブシステム1の更 報告し、遠隔地の配像サブシステム2におけるデータの 更新 (反映) が、近い方の記憶サプシステム1における その指示対象が遠隔地に設置された記憶サブシステム2 ステム1に、データの更新(費込み)指示が有った場合、 処理とは非同期に実行される処理手順をいう。 2

伝送時間、格納処理時間等により、ホスト1の処理が待 たされることがない。ここで、遠隔地とは、いわゆるト 【0032】このため、近い方の記憶サブシステム1で 必要とされる処理時間でデータ更新が終了するので、遠 隔地の記憶サブシステム2へのデータの格納に起因する ランスコンチネンタルネットワークと称される、附近地 より違いが、距離の制約なく通信又はデータ転送可能な 地点を買うものとする。

25

は、ホスト1から受領した更新データブロックを記憶サ テム1は、自己のスケジュールで、ホスト1の処理とは 【0033】より具体的には、非同期リモートコピーで プシステム1で受飯し (■)、ホスド1に対し更新デー タブロックの掛き込み完了を行う(■)。 記憶サブシス 非同期に、記憶サブシステム2ヘデータを転送する。 2

【0034】 強隔地又は附近地へのデータ転送経路の複 **単化、途中のデータ転送経路のポトルネックにより、デ** 一タ転送中の当該データの順序性は保証されない(図 6、点線の楕円内参照)。 35

いてデータを転送する場合がある。また、転送先まで遠 する交換機、ルーターその他の通信中継機器により、転 のように複数の転送経路を用いてデータを転送する場合 早い経路とを介してデータが送られるため、転送先にお 多くは高速転送のため、転送元から複数の転送経路を用 距離となると、転送元は1つの転送経路であっても介在 には、経路によっては時間的な差異が生じ、遅い経路と 送先まで転送経路が1本であることは保証されない。こ 【0035】一般に、データ転送の性能を上げるため、 いてデータの順序が保存されないのである。 45

【0036】図6の楕円内に一例を示すが、データ転送 経路上の順序を、Data#1、Data#2、Dat

a#4、Data#3としている。記憶サブシステム2 て正規の順序に並べ直しているからである。この更新処 スやジャーナルファイルシステムは回復処理を行うこと ができる。逆に、更新処理の直前に災害が発生したとき は回復処理は不可能であるが、ホストへの応答とは非同 期に、記憶サブシステム国で連続的に固断無くデータ転 こおける更新順序はData#1、Data#2、Da ta#3、Data#4の順序である。記憶サブシステ ム2において、転送されてきたデータの順序をソートし **閏の直後に不慮の災害が発生しても、データ更新の順序** が守られているため、記憶サブシステム2のデータベー 送処理を行うことで、データ不整合を極小化でき、巨視 的に見て、常時、更新データの順序性を確保できる。

ホストからのデータ更新順序を示すシーケンス番号情報 このような一連のデータ転送・処理に必要な処理時間の 後は、データの順序性が遠隔地の記憶サブシステム2に 【0037】本実施の態様では、ホスト1からデータブ をデータに付して転送している。このため、配像サブシ ステム2で、シーケンス番号情報に基づくソート制御を これに固有なデータ処理を連続して行うこと(非同期リ モートコピー)で、常時、データ更新の順序性を保証す ロックを受領し、記憶サプシステム2へ転送する際に、 行い、順序性を保証して、データの格納を完了できる。 おいて保持されている。このように非同期のコピーを、 ることができる。

が保証されるため、広域データストレージシステムの利 [0038] 非同期リモートコピーは、ホスト1の処理 生能を落とさず、記憶サプシステム 1 及び 2 の間の距離 を拡大できる特長があり、かつ、常時、データの順序性 用者が自己の業務を遂行する上で、ほぼ任意の時点のデ **遠隔地に股団された記憶サブシステムにおいて確保でき ータベースやジャーナルファイルシステムの整合性を、** る特長を有している。

20

>図1に本発明の広域データストレージシステムの全体 構成を示す。図9は本発明の別の広域データストレージ システムの全体構成を示す図である。図10は図1と図 [0039] <広域データストレージシステム、その1 9の構成の組み合わせによる応用例を示す図である。

サブシステムを設置する。各データセンタには複数の記 プリケーションはデータセンタ1に接続されたホストで [0040] 図1において3ヵ所のデータセンタに記憶 **憶サブシステムが設置されても良いし、それらがリモー** トコピー機能を伴った接続形態となっていても良い。ア 実行される。尚、ホストとデータセンタ1とのデータ転 送紙路は、ファイバーチャネル、メインフレームインタ フェース、イーサネット (登録商標) LAN、公衆回 は、インターネットその他専用回線である。

【0041】データセンタ1とデータセンタ2は附近地 る。データセンタ1とデータセンタ3は盗隔地に存在 に存在し、同期転送によりデータ転送し得る構成であ

いれらの回は非回避リモートコピーの技術によりデ **ータ転送し得る構成である**,

[0042] 正常な迎用形態では、ホストからデータセ ンタ1が受領した更新データは、データセンタ1に設置 された配像サブシステムに格納され迎用される。この更 新データは、附近地に設置されたデータセンタ 2 の配像 サブシステムへ、ファイバーチャネル、メインフレーム インタフェース、イーサネットLAN、公衆回級、イン ターネットその他専用回線を介して、同期転送される。 9

つまり、データセンタ1とデータセンタ2では、配位サ プシステム間のデータ整合性は巨視的には絶えず保たれ

2

タセンタ3、それぞれの間のデータ転送経路は同一種類。 [0043] 正常な運用形態では、また、ホストからデ ータセンタ1が受倒した上記の更新データを、遠隔地に 数回されたデータセンタ3の記憶サブシステムへ、上獄 と同様な専用回線を介して、上紀の同期転送処理と同時 に、非回期リモートコピーの技術で伝送される。他、デ **ータセンタ1とデータセンタ2、データセンタ1とデー** の回線にする必要はない。 20

【0044】 データセンタ 1 とデータセンタ 3 との間は 遠距離であり、この間の転送に起因する更新データの到 **労順序の不整合が生じる。また、転送元となるデータセ** ンタ1の記憶サブシステムには、転送先で未反映のデー **タとなる差分データが、存在することとなる。しかし、** 52

本発明の非同期リモートコピーでは、所定の非同期転送 に固有のデータ処理後は、データベースやファイルシス テムの回復処理に必要な、ホストからのデータ順序性を **保証しているために、不整合を生じていたデータの順序** タ1とデータセンタ3の記憶サプシステム間では、上記 を回復させることが可能である。この結果、データセン ホストから受倒した更新データの順序性は似たれる。

万一のリカパリ処理に備え、データを転送する通信模路 は敷設・準備されているが、このストレージシステムの 正常な運用時にはホストからの更新データは転送されな [0045] データセンタ2とデータセンタ3の間は、

い。データセンタ1での災害・障害発生の際に備え、正 常な運用形態で、データ転送の進捗状態を則合せるコマ ンドが、この通信額路を介して、データセンタ2からデ **一タセンタ3へ、又は逆にデータセンタ3からデータセ** ンタ2へ、送受信されることとなる。尚、敷散・坤鰡さ れた通信線路は、ファイパーチャネル、メインフレーム インタフェース、イーサネットLAN、公衆回教、イン 35 ç

れたホストからの更新データの到益を、配位サブンステ 【0046】正常時には、配像サブシステム1と配像サ プシステム3との間の非同期リモートコピーにより為さ ム2から発せられる。データ進渉問合せコマンド。によ ターネットその他専用回線である。 45

り、データセンタ2とデータセンタ3の間の通信級路を 介して、問合せる。

20

20

【0047】"データ進捗問合せコマンド"の起動は、記憶サプシステム2のスケジュールに従って為される。記 節サブシステム 1からの同期転送によるデータの受領の の時間間隔で纏めて問合せても良い。所定の時間開隔と 問合せても良いが、後述する転送状態/ビットマップの 管理:これに基づく差分データの管理に時間が費やされ 過ぎない程度となる。尚、1回の問合せで複数のビット タイミングで、当該コマンドを発行しても良いし、所定 しては、例えば100msecから500msec年に マップを検査するようにしても良い。

記憶サブシステム3との間で、直接、データの転送は行 【0049】 万一、データセンタ1で障害が発生したと きには、データセンタ2のホストを用いて、これまでの サブシステムへ切り替えることをいい、古くは、ホット [0048] 正常な運用時には、記憶サブシステム2と 記憶サブシステム 2 と記憶サブシステム 3 との間の差分 データを、リカバリ処理に備えて敷設されたデータ転送 の通信線路を用いてデータセンタ2からデータセンタ3 尚、フェールオーバーとは、プライマリーシステムから へ転送する。差分データのみの転送で即時に広域データ なわれない。このため、記憶サブシステム2が"データ **進捗問合せコマンド"を発行して、記憶サブシステム1** システム運用を総行し(ホストのフェールオーバー)、 と記憶サブシステム3のデータ更新状況を把握する。 ストレージシステムを回復させることが可能である。 スタンパイとも呼ばれていた。

旧に伴う、データセンタ2とデータセンタ1との間の同 **ジシステムを復旧させることがでる。但し、障害発生前** 後で、データセンタ1とデータセンタ2の役割が入れ替 【0050】この後、データセンタ2からデータセンタ モートコピーを行うこととすれば、データセンタ1の復 切転送の復旧により、障害発生前の広域データストレー 3へ、上記の通信線路を用いて、上述のような非同期リ

は、附近地に存在する、相互に同期転送により連結され タセンタと、遠隔地に存在する2つのデータセンタとを コピーの技術で連結する広域データストレージシステム [0051] このように、附近地に存在する2つのデー 統合し合計3つのデータセンタとすることで、リモート とする。こうしておけば、中小規模の災害・障害のとき る。2つのデータセンタの記憶サブシステムのデータは 同期転送により巨視的にみて一致しており、フェールオ たデータセンタの一方で他方の代替を行うことができ ーバーが即時に行なえるからである。

間のデータ転送経路を選択する場合には、復旧後は、広 [0052] <広域データストレージシステム、その2 >図1のデータセンタ2とデータセンタ3との間の通信 線路が非常用であるため、この通信線路を選択せず、障 お・災害復旧後のデータセンタ1とデータセンタ3との 壌データストレージシステムは、図9の構成となる。

システム3が非同期リモートコピーで、それぞれ、接続 【0053】図9は、記憶サブシステム1と記憶サブシ ステム2が同期転送で、記憶サブシステム2と記憶サブ された例である。図1の広域データストレージシステム において、データセンタ 1 からデータセンタ 2 へ運用を 切り替え、データセンタ2を主たる運用サイトとし、災 書・障害復旧後は、データセンタ2からデータセンタ1 ヘデータを同期転送させる一方で、データセンタ1から データセンタ3ヘデータを非同期転送させる構成となる からいある。 05

[0054] 図9の場合において、直接データ転送に関 へ、" データ転送進渉問い合せ" コマンドが発行され、デ ータセンタ 3 が応答して結果をデータセンタ 1 へ返す構 成となっている。また図10は、図1と図9を組み合せ た構成である。配位サブシステム3と5との間、記憶サ プシステム2と5との間が、"データ進渉問い合せコマ 与しない記憶サブシステム1から記憶サブシステム3 ンド"の発行・応答の経路に該当する。

【0055】上記の広域データストレージシステムの態 様であれば、大規模な災害や、附近地に存在する2つの とで、災害直前のシステムが運用してきたデータを引き データセンタ 3 のホストヘフェールオーバーするこ 継いて処理でき、また、データの喪失を最小限度とする データセンタに相次いで除害が発生した場合であって ことができる。 25

が全滅する程度の災害が発生したときは、遠隔地に存在 するデータセンタ3又は5(図1,図9,図10)の記 **箘サブシステムを生かすことができる。ホストからの更 一が行なわれているからである。但し、災害による未反** [0056] つまり、附近地にある2つのデータセンタ 新データの順序性が確保されつつ、非同期リモートコピ 20

1、図9及び図10では、同期転送によるコピー及び非 [0057] <記憶サブシステムの構成について>図 吹のデータは復旧できない。

は、1個の論理ポリュームに対するデーク受領を、同期 転送し、更に非同期転送して、附近地と遠隔地の双方に イクロコードで実現される機能である。ホストや別の記 リモートコピーは、1 論理ポリュームと1 論理ポリュー 【0058】これらは記憶サブシステムの制御装置のマ 憶サブシステムからの更新データは、一旦、キャッシュ リモートコピー機能でデータ送信制御を行なっている。 同期リモートコピーの組み合わせを示している。本来、 ムをデータ転送技術で結合したものである。本発明で 5 (図2) に格納される。この時点では、当該データ 35 40

は、まだ記憶サブシステム内のハードディスクドライブ にRAID制御により書き込まれていない。キャッシュ 5 内で当核データの転送制御情報を加え、別の記憶サブ システムヘリモートコピー転送したり、複数の記憶サブ システムとのリモートコピー構成を同時に実現する制御 を行う。同期転送と非同期転送による組合せを守ること 45 20

ファイルシステムがリカバリ可能な論理ポリュームを常 データの更新順序を保った、データベースやジャーナル こより、いつ災害が発生しても、各データセンタでは、

[0059] 図2は、記憶サブシステムの蝦略構成を示 時、保持していることとなる。

す図である。

の接続先とデータの送受を行うチャネルアダプタ3、デ インタフェース8 (ディスク1/F8) を介して制御す [0060] 制御装置1は、ホスト及びリモートコピー イスク装置 2 内のハードディスクドライブ 7 をディスク るディスクアダプタ9を有する。

接続されている。尚、バス構成は一例であり、必要に応 複数設けてクラスタ構成とし、複数の制御装置1を連絡 [0061] チャネルアダプタ3とディスクアダプタ9 は、それぞれ、マイクロプロセッサを有し、データ転送 パス11・制御パス11を介してキャッシュメモリ5と じてクロスパー構成としても良い。また、制御装置1を する共通の第3のパスを追加しても良い。

る。制御情報、構成管理情報、転送状態/ビットマップ 【0062】ホストとの間や、リモートコピーの接続先 とデータ送受を行う際の格納元は、キャッシュ5であ は、制御メモリ6に格納されている。

ルアダプタを分けて搭載している。ホストから受倒した [0063] リモートコピーには送信及び受信の機能が あり、本実施例ではホストから1/0を受領するチャネ 1/0は、一旦、キャッシュ5へ格納される。リモート コピーの転送先情報や後述する状態管理/ピットマップ は、制御データとして制御メモリ6に格納され、マイク ロコードにより制御される。

[0064] キャッシュに受領したデータは、ディスク アダプタ9によりハードディスクドライブ7へRAID 制御で書き込まれる。これとは別の処理である、マイク ロコードを用いた制御により、予め定義されたリモート コピー転送先への送信制御が行なわれる。

33

[0065] 例えば、ホストから受領したデータが、後 統するリモートコピーの対象であり、非同期転送による データ送信を行うと定義されていた場合には、キャッシ ュ5の内部のデータに対して、データ受領順にシーケン ス番号を付与する。これはデーク更新を示す I D情報で もある。シーケンス番号を付与されたデータは、チャネ ルアダプタ3のリモートコピー送倡機能により、当核シ 一ケンス番号と共に送信される。

[0066] 別の実施例で、ホストから受領した更新プ に、非同期転送用にも加工され、シーケンス番号が付与 ロックを、複数の論理ポリュームと接続するリモートコ ピー制御が定義されていた場合には、キャッシュメモリ 5の内部のデータは、同期転送用に加工されると同時 されて、それぞれ、チャネルアダプタ3で附近地又は、 遠隔地に向けて送信される

20 [0067] 図2は本発明を実現する一例であり、本発

明はハードウエア構成に依存しない。リモートコピー接 統が記憶サブシステム間で実現可能であれば、マイクロ プロセッサによる論理的なサポート、マイクロコード側

თ

0

ີດ 2 2

ı က

特開200

リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

[0068] <転送状態/ピットマップ>図4は、転送 記憶サブシステムの内部に、災害・障害の復旧の際にペ る。)の一例を示したものである。これは、直接データ **転送を行なっていない2つのデータセンタに設置された** 伏態/ビットマップ(適宜、ビットマップと略記す 05

記憶サブシステム)のデータ更新の進捗状況を知るため タ2とデータセンタ3との間で、非常時のためにペアが ム5、記憶サブシステム3と記憶サブシステム5との問 アを組むであろう相手(別のデータセンタに設置された に用意されたものである。例えば、図1ではデータセン 組まれる。図9の広域データストレージシステムであれ で、図10では、記憶サブシステム2と記憶サブシステ で、それぞれ、非常時のために、ペアが組まれることと ば、記憶サプシステム1と記憶サプシステム3との問 2

【0069】転送状態/ビットマップは、ペア(対)と 間の論理ポリュームの実体に対し、2個以上の転送状態 /ピットマップを持ち得る。各ピットマップは、ペアや ペアとなる場合を想定した定義付けにより、相手の論理 ポリュームとの差分管理を行うために使われる。ビット マップの中のプロックナンバは、艶型ボリュームの更新 なる論理ポリュームに対して必要であり、本発明では1 を管理する最小単位であるプロックに対応させた番号で 20 52

【0070】ホスト1/0は、このブロックナンバと同・ 可変である。一方、ピットマップは、50kB弱の大き kBから1000kB程度まで種々の大きさがある。ホ 常、最小で512パイトとされ上限も設けられているが さ、又は700kB程度の大きさのものもあるが、20 スト1/0の1プロックに対して、必ずしも、1ビット 一単位である必要はない。ホスト1/0の単位は、通

35

【0071】ブロックナンバに対応するブロックの内容 が更新されれば、差分管理は当該プロックナンバ全体と なり、同期(リシンク)を行うときに当該プロックナン バのデータ全体が転送されることとなる。 マップが対応する訳ではない。

数論理ポリュームの更新された単位として、リモートコ 当該更新されたプロックのみを転送する目的で、相手論 【0072】ビットマップは、ブロックナンバ毎に、当 ピーによるペアを再構築する際(再同期、リシンク)に 理ポリュームに転送すべき"Update"情報を持つ。 40 45

1)、であれば転送対象であることを示す。通常のUp dateは、ホストからのコマンド単位で為されること から、カウンタ値が0であることに基づき、Updat つまりUpdateフラグがOn (図4の実施例では eフラグを0とする。

[0073] ピットマップは更に、同一ブロックナンバ で、複数回の更新を配録するカウンタ値を持つ。カウン きさが、ホストから更新されるデータブロックより大き い場合には、このカウンタ値を使うことにより確実に相 夕値は、更新が無ければ"0"、3回更新されれば"3"と なる。プロックナンバーで表されるデータブロックの大 手論理ポリュームへ更新データのみを転送できることと

2 格納されたプロックナンパとカウンタ値と、問合先の記 愉サプシステムのピットマップのプロックナンバとカウ 大きい場合に、所定の記憶サブシステムのビットマップ [0074]後述の"データ進捗問合せコマンド"の中に ンタ値との比較を、データコピー監視機能(後述)で行 う。この際に、ある記憶サブシステムが持つカウンタ値 が、この記憶サブシステムに送付されて来た"データ進 **捗問合せコマンド"に記述されたカウンタ値と等しいか** のカウンタ値は1域算される処理を受ける。

位サブシステムのピットマップのカウンタ値は何ら処理 ド"に記述されたカウンタ値未満である場合は、その記 を受けない。そして減算したかしないかを、" データ進 [0075] 送付されて来た。データ進捗問合せコマン **歩岡合せコマンド"に応答して返す。**

その記憶サプシステムにおいて格納済み、即き込み済み であることを意味する。また"未満"の場合には、データ 【0076】その記憶サブシステムのピットマップのカ ド"に記述されたカウンタ値"以上"の場合には、データ ウンタ値が、送付されて来た。 データ進捗問合せコマン **又新の進捗は、正常なリモートコピー機能により既に、** が未到着であることを意味している。

1パイト分をカウンタ値として割り当てた場合には、2 ウンタ値のUpを行わず、Updateフラグを恒久的 グの解除(0を入力すること)は、相手論型ポリューム 56回を超える管理はできない。この例では同一プロッ クが256回を超えた更新を受けた場合には、最早、カ [0078] このような恒久的な指定がなされると(図 への転送が完了しコピーが確定したことを、このピット マップを有する記憶サブシステムが認識するときまで行 る、恒久的指定のなされたプロックのUpdateフラ 【0077】図4のカウンタ値は有限であり、例えば、 に立ててしまう処理を行う。つまり図4でカウンタ値 4、Over Flow)、ピットマップで特定され に、Over Flow、を意味する情報を格納する。

【0079】カウンタ値を用いた更新管理を行う理由を 次に補足説明する。

クに対応させてピットマップ管理を行うのは、災害・障 [0080] 例えば、50kB程度のデータ肌を有する に、この50 k Bのデータのうち、異なる3 箇所が、異 なる時刻において、それぞれ更新されたとする。トラッ トラックに対応させてピットマップの管理を行う場合

杏後の仏旧 (再同期、リシンク) において扱う単位が ラック単位であるためである。

る時刻でUpdateフラグが1であることのみを確認 しても、その後の時刻に2度目、3度目の更新があった [0081] カウンタ値による管理を行なわない場合に は、Updateフラグのみ監視することとなるが、あ 場合には、2度目以降のデータ更新を見逃してしまう。 新たにカウンタ値の概念を導入して、ホストからのコマ ンド単位で為される同一データブロック(ここではトラ ックの一部)の更新を敬細に監視することで、かかる不 都合を防ぐことができる。

[0082] 次に、図2の制御装置1の内部でマイクロ コードにより実現される転送状態/ビットマップの機能 について定義する。論理ポリュームはリモートコピーの 対となる論理ポリュームとの間で下記の転送状態を有す る。これらは同期転送又は非同期転送に依存しない。

性を保証して、双方のポリューム間で、同一のデータを 【0083】1) 「正常ペア状態」とは、データの順序 2 瓜に保持している状態をいう。

とは、データの更新をビットマップに登録する状態をい 【0084】2) 「転送抑止ピットマップ登録の状態」 う。未だペアの相手へデータの転送は行なわれていな 20

【0085】3) 「ピットマップ使用のコピー状態」と は、「転送物止ピットマップ登録の状態」から「正常へ ア状態」への移行期をいう。2m化のためのコピーの初 明状態に当たる。 [0086] 4) 「障害状態」とは、障害によりデータ を転送できない状態をいう。ピットマップに登録され

[0087] 5) 「ペア無ピットマップ登録状態」と は、本発明固有の特殊な状態をいう。災害・障害前に、 30

相互にデータ更新状態を監視し保持する必要から生じた 【0088】6) 「ペア無状態」とは、ピットマップは 32 用意されているが、未だペアを組んでおらず、データ更

新の惰報が登録されていない状態をいう。

く、" 転送拘止ビットマップ登録の状態"というサスペン ド(Suspend)状態で兼ねても良い。ここで、サ 【0089】「ペア無ビットマップ登録状態」が存在す ることが本発明の特徴となる。この状態を持つことな

スペンド状態とは、論理ポリュームへのデータの更新状 旗を、ピットマップでのみ管理し、リモートコピーによ 【0090】「ペア無ピットマップ登録状態」を持つの る転送制御を行なわない状態をいう。 45

は、転送状態/ピットマップをペアで持つ必要からであ る (図3)。 例えば、図1の広域データストレージシス [0091] データセンタ3が保持するデータを監視す テムにおいては次の理由による。

リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

るため、データセンタ3の記憶サプシステムの内部の論 理ポリュームに対応して設けられた転送状態/ピットマ ップに、データセンタ2のデータ更新状態を持つ必要が 理ポリュームに対応して設けられた転送状態/ピットマ ップに、データセンタ3のデータ更新状態を持つ必要が あり、且つ、データセンタ2が保持するデータを監視す

ンタ1とデータセンタ3のリモートコピーの差分管理情 **喜が発生しても、状態把握が可能で、ピットマップによ** る未転送データブロックの記憶と、障害回復後に更新部 [0092] 図9の広域データストレージシステムにお **ハでは、データセンタ2の障害発生に備えて、データセ 報から、データセンタ1とデータセンタ3との間でペア** 構築を目的として、「ペア無ピットマップ登録状態」を の結果、記憶サブシステムやデータ転送経路のどこに障 データセンタ 1とデータセンタ 3で持つ必要がある。 分のみの差分転送が可能となる。

ードの制御により自由に実装できる。例えば、マイクロ [0093] 転送状態/ピットマップの機能は、上記の 様な制御を実現するマイクロコード及びピットマップと ば、図2のマイクロプロセッサ4のマイクロコードと餓 **脚メモリ6で行なわれるが、先に示した様にマイクロコ** 関連する制御テーブルから成る。具体的機能は、例え

ドを記憶サブシステム2から記憶サブシステム3へ送信 する。例えば、記憶サブシステム1の障害の際、実際の **並分データの転送に腐しては、配像サブシステム2と配** について、論理演算を行う。その結果に基づき、相当す 【0094】 <広域データストレージシステムの巡用> 図3は、図1の広域データストレージシステムが正常に 運用されている場合の基本的な制御方法を説明するため の蝦略図である。正常巡転ではデータ進捗問合せコマン 位サプシステム3との間で、転送状態/ピットマップの 機能を使用し、両方の記憶サブシステムのビットマップ るデータプロックのみを記憶サブシステム2から記憶サ ブシステム3へ転送している。図8に、図1の広域デー タストレージシステムのデータセンタ 1 に降唐・災害が 発生した場合において、非同期リモートコピーを再開さ せる蝦略の手顧を示す。

【0095】図8において、正常な運用では、データセ ンタ1から附近地のデータセンタ2へ同期転送によりデ 3へは非同期転送によりデータの更新順序を確保したコ ピーが行なわれている。そして、データセンタ2の記憶 サブシステム2のスケジュールで、データ進捗問合せコ 2と3とは管理情報を避り取りして、データの差分管理 **ータの二重化が図られる一方で、遠隔地のデータセンタ** マンドがデータセンタ3に対し発行され、データセンタ

と、データセンタ2の配像サブシステムは、非同期転送 に、データセンタ2と遠隔地のデータセンタ3によるシ により、並分データをデータセンタ3へ送付し、即時 ステム運用を回復できる。

らのピットマップを用いた機能を有する。配位サブシス 対し、転送状態/ピットマップ#1に対応する機能及び は、1輪里ボリューム当り2個符ち、それぞれが、これ デム1は、記憶サブシステム2と記憶サブシステム3に [0091] 図3において、転送状態/ピットマップ ピットマップ#2に対応する機能を持つ。 9 9

[0098] 記憶サブシステム2と記憶サブシステム3 ピットマップ#3及び#6の機能をそれぞれ持つ。これ ら#1と#3、#2と#6のそれぞれの機能は、正常週 は、同期転送及び非同期転送の各々について転送状態/ 気の際には、"正常ペア状態"を格納している。

ムが正常に運用されているときには、転送状態/ピット [0099] 転送状態/ピットマップ # 4 及び # 5 の機 テム3が恃っている。この広域データストレージシステ 能は、それぞれ、記憶サブシステム2及び配億サブシス マップ#4及び#5の機能は、上述の。 ペア無ピットマ ップ登録、状態を保持する。 20

[0100] 転送状態/ピットマップ#4の機能は、記 ステム2の論理ポリュームに対する遵分管理を、それぞ を、転送状態/ピットマップ#5の機能は、記憶サブシ 佐サブシステム3の論理ポリュームに対する差分管理 れ行う。

52

プロセッサ10による実現も可能である。またマイクロ

プロセッサが1台のみの制御装置でも実現できる。

[0101] 図10の拡張として、ホストからの1/0 を受倒する、第1のデータセンタに設置された配塩サブ システムの値御装置1が、N台の同期転送のコピー先

と、M台の非同期リモートコピーのコピー先を持つ構成 では、その制御装置1は、N+M個の転送状態/ピット マップの機能を有する。これに対応する、選隔地又は附 近地の配憶サブシステム(コピー先)も、転送状態/ピ ットマップを持つこととなる。この結果、側御装置1や データ転送経路のどこに陞書が発生しても、状態把揺が 可能で、ピットマップによる未転送データブロックの記 億と、災害回復の際の更新部分のみの差分転送が可能と 20

[0102] <データコピー監視機能>次に、データコ パー脂褐繊細にしいた説明する。この微幅には、パット マップの制御機能、リモートコピーのステータス管型機 能、構成管理機能、データ連捗問合せコマンドの制御機 能、リモートコピーのデータ転送指示機能等が含まれ 9

同期転送によるデータブロックを配位サブシステム1か ら受倒する。 本データは配像サブシステム2のキャッシ ュメモリに格納されディスクドライブで配修される。こ の際、転送状態/ピットマップ#4の機能により当骸デ **一タブロックが登録される。図4のビットマップに登録** [0103] 図3の記憶サブシステム2の制御装置で、 20

るため、データセンタ2の記憶サブシステムの内部の論

20

[0096] データセンタ1に災害・障害が発生する

【0104】次に当該プロックナンバとカウンタ値を格 **附した" データ進渉問い合せ" コマンドを配憶サブシステ** ム2から記憶サブシステム3に対して発行する。発行の タイミングは同期転送に基づいても良いし、記憶サブシ

【0105】記憶サプシステム3の制御装置で、記憶サ ブシステム 2からの" データ進渉問い合せ" コマンドを受 カウンタ値を切り出し、記憶サブシステム3の該当する **倒し、転送状態/ビットマップ#4のブロックナンバと** ステム2の独自のスケジュールで行なっても良い。 転送状態/ピットマップ#5のそれらと比較する。

ば、同期転送に係るデータと、非同期リモートコピーに [0106] その結果、転送状態/ピットマップ#5の 係るデータとが一致しているので、転送状態/ビットマ ップ#6の対応するプロックナンバから、カウンタ値を し、かつ、カウンタ値が転送されて来たもの以上であれ プロックナンバがUpdateフラグ1 (更新) を示

合には、Updateフラグを"0"とする。カウン夕値 [0107] 減算の結果、カウンタ値が"0"となった場 が"Over Flow"である場合には、何も操作しな

[0.108] また、転送状態/ピットマップ#5に登録 れなかった場合には、#5への更新は行なわず、これを り、Updateフラグが"0" (Off) で更新が示さ データ進捗問合せコマンドの結果として記憶サブシステ されていたカウンタ値が、記憶サプシステム2からの問 合せコマンドから抽出されたカウンタ値未満であった

52

#6の転送状態/ビットマップのカウンタ値を減算する 記憶サプシステム1から記憶サプシステム3へ非同期転 ということは、記憶サブシステム1から同期転送により [0109] #5の転送状態/ビットマップの機能が、 **既に記憶サブシステム2へ到着したデータブロックが、** 送により到着済であったことを意味している。

【0110】データコピー監視機能は、本応答結果を用 いて記憶サブシステム2の転送状態/ビットマップ機能 には、記憶サブシステム2の制御装置でも転送状態/ビ ットマップの機能でカウンタ値の減算、Updateフ の制御を行なう。配像サプシステム3で"データ進渉問 い合せ"コマンドのブロックナンバとカウンタ値が既に 登録されていた旨の応答を返す場合(減算できた場合) ラグの操作を同様に行う。

ば、記憶サプシステム1から記憶サプシステム3へのデ ータの非同期転送が未完であるとして、配盤サブシステ ム2の転送状態/ピットマップ#4の機能は、自己のピ [0111] 当該コマンドの応答結果が、未登録であれ ットマップに更新状況を保持する。これは後に更新差分 那分のみを再同期させる際の対象となる。

【0112】この時点で記憶サブシステム1が重大障害

を持ち、記憶サブシステム2と記憶サブシステム3との しなければならない場合には、ビットマップを参照した 結果、未転送のデータのみ、即ち、差分のデータブロッ 間でリモートコピー構成を再構築(再同期、リシンク) クのみを、記憶サブシステム2から3へ転送すれば良

ア"を構築できる。これを実現する機能を"データコピー い。その結果、差分データの転送だけで即時に"正常ペ 監視機能"と呼ぶ。

>図9の広域データストレージシステムにおいて、記憶 [0113] <正常な運用の際に、直接データ転送を行 なわない記憶サブシステム間での差分管理方法、その1 サブシステム2に障害が発生した場合に、記憶サブシス テム1と記憶サブシステム3との間で非同期リモートコ ピーによるシステム運用の復旧を図るときを考える。

サプシステム2の制御装置1の論理ポリュームに同期転 【0114】このために、ホストからデータ更新を受領 した記憶サブシステム1の制御装置1(図2)は、記憶 送のコピーによるデータ転送を行う際に次の処理を行な [0115] 転送するブッロクの位置情報を、記憶サブ システム 1の制御装置 1 に存在するビットマップに、記 る。このとき既に転送したブッロクが記憶サブシステム 3において更新されていたときは、ピットマップのカウ 憶サプシステム 3 の論理ポリュームの更新怙報を格納す ンタ値を1増加 (インクリメント) する。 20

[0116] 記憶サブシステム1の制御装置1は、記憶 サブシステム2の制御装置1に対して同期転送が完了し た後、記憶サブシステム3の制御装置1に対して、同期 転送したデータブッロクが、記憶サブシステム2の制御 装置 1 を経由して到着したか否かを問合せるため、記憶 サブシステム 1 と記憶サブシステム 3 とを結ぶ通信線路 を用いて、確認コマンドを発行する。 30

【0117】確認コマンドには、ホストから受領した更 新データの記憶サブシステムにおけるデータブロックの ンドを受領した記憶サブシステム3の制御装置1は、記 ブロックナンバとカウンタ値が含まれている。確認コマ タブロックが、確認コマンドで問合されたブロックとー **躗サブシステム2の制御装置1経由で既に存在するラ** 致するか否かを判定する。

[0118] 記憶サブシステム3の制御装置1は、記憶 サブシステム2の制御装置1の論理ポリュームに対する 転送状態/ビットマップの機能の他に、記憶サブシステ ム1の制御装置1の論理ポリュームに対する状態管理/ ビットマップの機能を持つ。 40

[0119] 記憶サブシステム3の制御装置1は、記憶 サブシステム2の制御装置からデータを受領すると、記 億サブシステム1の制御装置1の状態を格納すべく、自 己の持つ転送状態/ビットマップへ登録する。このビッ トマップでは、論理ポリューム内のアドレスに関わるプ ロック位置に対する更新情報を有し、さらに複数回の同 45 20

一プロックに対する更新を管理するためにカウンタ値を

上の場合には、データの到着が正常に完了していると判 ム1の制御装置1から発行された確認コマンドのブロッ 致又はカウンタ値が確認コマンドにあったカウンタ値以 【0120】記憶サブシステム3の制御装置1の転送状 態/ビットマップに登録された結果は、記憶サブシステ 断し、転送状態/ピットマップの機能を用いてピットマ クナンバ及びカウンタ値と比較される。比較の結果、 ップのカウンタを1減算する。

は、記憶サプシステム3の制御装置1から返される結果 が、記憶サプシステム3ヘデータプロックが記憶サブシ 上述の記憶サブシステム3の制御装置が為したように、 転送状態/ビットマップの機能を用いてビットマップの ステム2を経由して到着していることを示す場合には、 [0121] 他方、記憶サブシステム1の制御装置1

5記憶サブシステム1と、記憶サブシステム2の内容を. **舂を持ち、同期及び非同期転送によるデータ送受が行な** 非同期リモートコピーにより格納した記憶サブシステム 【0122】以上のように、ピットマップを監視・管理 **「ることで、記憶サブシステム2が災害等により重大障** 3 との間で、非同期リモートコピーを構成することがで えなくなった場合であっても、ホストが1/0を発行す

[0123] この際に記憶サプシステム1と3の、それ 差分データのプロックのみを転送することにより、即時 り、論理ボリュームの全データをコピーすることなく、 ぞれの制御装置の転送状態/ビットマップの機能によ

有する対応する値との比較を行う。

テム1と2並びに記憶サブシステム1と3、それぞれの なわない記憶サブシステム間での差分管理方法、その2 [0124] <正常な運用の際に、直接データ転送を行 >図1の広域データストレージシステムにおいて、ペア になっている論理ポリューム間、つまり、記憶サブシス 間のデータ更新状態の管理のために、転送状態/ピット マップの機能が各論理ポリューム年に用意される。

と3のそれぞれの制御装置1の間で、先ず差分データを [0125] 記憶サブシステム1の制御装置1で障害が 発生し、同期転送のコピー及び非同期リモートコピーの 双方が継続不可となった場合には、記憶サブシステム2 コピーし両者を一致させる。次いで、記憶サブシステム 2と3の間で非同期リモートコピーを構成する。

システム3の制御装置1の配下の論理ポリュームの管理 【0126】ホストから更新すべきデータを受領した記 **箆サブシステム1の制御装置1は、記憶サブシステム2** 5. 記憶サブシステム2の制御装置1は、受領したデー タブロックの位置情報 (ブロックナンパ) を、記憶サブ の制御装置 1 へ同期転送によりデータプロックを送出 し、これを記憶サブシステム2の制御装置1が受領す

トマップに格納する。転送状態/ビットマップは、受領 したデータブロックが更新された場合には、カウンタ値 を1増加 (インクリメント) する機能を備え、複数回の 情報との比較のために、自己が保持する転送状態/ビッ データブロックの更新を記録できる。

6

0

S

0

~

١

ი 0

特開20

リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

[0127] 記憶サプシステム2の制御装置1は、上記 て、データブロックが記憶サブシステム3へ到着したか の転送状態/ビットマップへ所定の管理情報を登録した 後、記憶サプシステム2の制御装置1と記憶サブシステ ム3の制御装置1との間を結ぶデータ転送経路を用い

否かを問合せる確認コマンドを、記憶サブシステム3の 【0128】確認コマンドは、記憶サブシステム2の制 制御装置1へ発行する。

御装置1が、同期転送により記憶サブシステム1から受 領したデータブロックの位置情報であるブロックナンバ と、データブロックが何回更新されたかを示すカウンタ [0129] 記憶サブシステム3の制御装置1は、記憶 サブシステム1の制御装置1から、非同期リモートコピ クナンバ) とカウンタ値を、記憶サブシステム2の制御 めに、自己の制御装置1が持つ転送状態/ビットマップ の機能を用いてピットマップに格納する。記憶サブシス テム3の制御装置1は、ピットマップと確認コマンドの 一の技術で受領したデータブロックの位置情報(ブロッ 装置1の配下の論理ポリュームの管理情報との比較のた

サブシステム3の制御装置1が持つ、記憶サブシステム= 【0130】 記憶サブシステム2から3へ問合せた確認 2の制御装置1の配下の論型ポリュームの管理情報であ コマンドが有するプロックナンバとカウンタ値と、記憶 又はカウンタ値が確認コマンドのカウンタ値より大きい る、これらの値とを比較して、確認コマンドの値と同一 場合には、転送状態/ピットマップの機能で、そのデー

【0131】減算した結果が0になる場合は、記憶サブ ビットマップの管理から削除する。上記の比較の結果が 一致しない場合には、配憶サプシステム3の制御装置1 システム2と3との差分データはないことになるので、 は、ピットマップのカウンタ値を操作しない。

タブロックのカウンタ値を1減算する。

33

ある判定結果を返す。この結果を記憶サプシステム2の 制御装置 1 が参照し、比較がカウンタ値を減算した場合 [0132] 記憶サブシステム3の制御装置1は、記憶 サブシステム2の制御装置1に、確認コマンドの応答で には、既に記憶サブシステム2と3の間で、同一のデー タブロックの更新が正常に終了していると断定する。

になる。 記憶サブシステム2の制御装置1は、自己の転 【0133】 記憶サブシステム3に핓新すべきデータブ のみ、更新に係るデータプロックが格納されていること ロックが届いていない場合には、記憶サブシステム2に 送状態/ビットマップの機能でこれを記憶する。 20

[0134] 記憶サプシステム2の制御装置1が、記憶 サブシステム3の制御装配1から確認コマンドの応答を クが未到着であった場合には、配億サブシステム2の制 御装図1が待つ、記憶サプシステム3の論理ポリューム の更新状態に対応する転送状態/ピットマップのカウン **更新に係るデータブロックが、記憶サブシステム2と3**

夕価は故算しない。このことは、そのピットマップは、

受倒し、記憶サブシステム3に更新すべきデータブロッ

リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

ム間で差分データのみをコピーし不一致をなくすことが プロックのカウンタ値を1減算する。カウンタ値が0の ときは、記憶サプシステム2と3との間で、更新に係る データブロックは同一であり不整合がないので、差分デ [0136] このように、正常週用の際に、直接データ が、災害・障害からの回復を想定して、論理ポリューム 間の整分データ管理を行っているため、配憶サブシステ **伝送を行っていない配像サプシステムの制御装配同士** ータのコピーの対象とはしない。 高速に行なえる。

は、上記の転送状態/ピットマップの更新に係るデータ

[0135] 他方、データの到着完了を示した場合に

との間で差分であることを示す。

図1に、図1の広域データストレージシステムが、フェ 場合の運用について簡単に説明する。 図3 で記憶サブシ ステム1に、図9で記憶サブシステム2に、図10で記 **位サプシステム1、2叉は4に、それぞれ重大な障害が** 超きた場合には、図7に示す様に、現存する2以上の配 **物サプシステムの阻で、リモートコピー構成の復婚を図** [0137] <フェールオーパー後のシステムの巡用> **ールオーバーにより状態を避移して図9の構成となった**

タのみコピーすれば、即時に、リモートコピーのペアを サブシステム2と3の間、図9の記憶サブシステム1と ステム3に対し、図9の構成では、記憶サプシステム1 から記憶サブシステム3に対し、それぞれ、記憶サブシ を要するコピー中に、再度、コピー元やデータ転送経路 システム1と記憶サブシステム3との間)で、差分デー [0139] 本発明を実施しない場合には、図3の記憶 3の間で、それぞれ、リモートコピー様成をつくるに軽 し、図3の構成では記憶サブシステム2から記憶サブシ ステムが保持するデータのフルコピーを行なわなければ ならない。大規模のデータセンタでは、コピーに長時間 を取し、リモートコピーの運用再開が遅くなる。長時間 に障害・災害が発生すると、データは破壊され喪失する 生成でき、リモートコピーの運用再開が可能である。

[0140] 図11を用いて、図9の構成におけるデー タコパー粗複繊循について簡単に説明する。 [0141] データ進捗問合セコマンドは記憶サブシス

テム1から記憶サブシステム3に対して発行される。デ ム2へホストから受領した更新データを転送した後、記 一監視機能"を作動させる。つまり、"データ進渉問い合 状態/ピットマップ#1と、配像サプシステム3の持つ 伝送状態/ピットマップ#3で、それぞれのUpdat 記憶サプシステム1が、同期転送により記憶サプシステ せ"コマンドを発行し、記憶サプシステム1の持つ転送 - タコピー監視機能は図1の場合と一部処理が異なる。 位サプシステム1から3に対し、上述した"データコピ e フラグ、カウンタ値を登録し、所定の操作を行う。

[0142] 記憶サブシステム1から3に、ホストから じデータが、記憶サブシステム3に届いたか否か、問合 記憶サブシステム1が受倒したデータ(トラック)と同 せた結果、未着であれば、記憶サブシステム1の転送状 塩/ビットマップ#1のビットマップは、そのまま保持 する。結果が到着であれば、つまり、#3のピットマッ ブのブロックナンバ、カウンタ値が同一であれば、Up dateフラグを削除し、#1のピットマップを削除す [0143] <再同期におけるその他の処理>データコ ピー監視機能で検出した"データ進渉問い合せ"コマンド の応答結果に、エラーや不具合(タイムアウト)が生じ たり、転送状態/ビットマップの機能に不具合が生じた 場合には、障害・災害の際に行われるべき回復処理に関 する差分管理を禁止する。 20

ピットマップは有限なカウンタ値の格納領域を有してい は、その後2以上の記憶サブシステム間で冗長度が維持 されていても、災害・障害発生後に再同期処理、差分コ [0145] 正常な運用において直接、データ転送を行 (確認コマンド送出) に対し、所定時間、応答が無い場 差分データのみ転送する処理を行なわず、禁止する。 ベ そのままペアの再度構築処理を行なわしめることは妥当 る。この有限値を超えて (オパーフロー)、同一データ [0144] 転送状態/ピットマップの機能において、 ピー処理が行なわれる際に、必ず更新対象として扱う。 る。非同期リモートコピーによるペアの再構築処理や、 アの相手のデータ更新状態を知ることができないため、 合は、タイムアウトであるとして再同期処理を禁止す ブロックが更新された場合には、そのデータブロック なわない記憶サブシステム間で選り取りされる問合せ

[0138] 本発明によれば、図7の様に、直接データ

阮送に関与していなかった論理ポリューム間(記憶サブ

50. る。以下、このようなデータの不整合を回避するための、 【0146】 〈非同期転送におけるデータの整合性の管 型>例えば、ホストが接続する記憶サブシステム1と記 位サブシステム2とが、記憶サブシステム1から記憶サ プシステム2にデータを複写する非同期転送で運用され ているとする。この場合、もし、記憶サブシステム1に おけるデータの哲き込み順と、記憶サプシステム2にお けるデータの掛き込み順とが異なると、両記憶サブシス テム1,2におけるデータの整合性が保証されなくな でないからである。

出組みについて説明する。

ック番号を割り当てる。そして、ホストからデータの低 が仍き込まれた場合には、図13に示すデータ管理情報 [0147] まず、各記憶サブシステム1,2における トごと)のプロックを区画して各プロックに固有のプロ き込みがあったプロックについて、そのプロック番号と データの魯き込み順に付与したシーケンス番号との対応 づけを制御メモリ6に管理する。例えば、図12に示す ように、プロック番号が56~59のプロックにデータ 記憶資源の記憶領域に所定サイズ (例えば、16Kバイ を削御メモリ6に作成する。

理情報を付帯させる。一方、これを受信した配位サブシ ステム2では、図15に示すように、データに付借して 送信されてきた前記データ管理情報を、傾御メモリ6に 管理情報、すなわち、シーケンス番号とプロック I Dの 組み合わせには、これに対応するデータのキャッシュメ モリ上の位置情報も対応づけて記憶されている。 記憶サ プシステム2は、前記データ管理情報のシーケンス番号 の顧番にこれに対応するキャッシュメモリ上の前配位置 【0148】 記憶サプシステム1から記憶サプシステム 2 への非同期転送に際しては、図14の転送データフォ **ーマットに示すように、転送するデータに前記データ管** 管理する。ここで制御メモリ6に管理される前配データ **作報に記憶されているデータを記憶資源に扱き込んでい**

[0149] 以上のようにして、ホストが記憶サブシス テム1の記憶資源に告き込んだ順番どおりに、記憶サブ 両記憶サブシステム1,2におけるデータの整合が保証 システム2の記憶資源においてもデータが你き込まれ、 されることになる。

プシステム2と、サイト3に設置された記憶サプシステ ム3とを備える。記憶サブシステム1には、この記憶サ る。記憶サプシステム1と記憶サプシステム3との問も 広域データストレージシステムは、サイト1に設置され た記憶サプシステム1と、サイト2に設置された記憶サ ブシステム1を記憶手段として利用するホストが接続す 【0150】<マルチホップ方式>図16 (a)·に示す 面信手段により接続される。

とは、記憶サプシステム1から記憶サプシステム2にデ **ータを複写する同期転送で運用されている。また、記憶** ステム2から記憶サプシステム3にデータを複写する非 る。なお、マルチホップ方式における各記憶サブシステ ム間の通信を同期転送とするか、非同期転送とするかは [0151] 記憶サブシステム1と記憶サブシステム2 サブシステム2と記憶サブシステム3とは、記憶サブシ 同期転送で適用されている。以下、このような形態のリ 任意に設定される。また、これら以外の転送方式であっ モートコピー制御方法を「マルチホップ方式」と称す

情すると (S121)、 母き込み対象データを自身の論則ボ [0153] 記憶サブシステム1は、ホストから铅き込 み対象データとその哲き込み要求 (Write 1/0) とを受 リューム (第1の記憶資廠) に告き込むとともに、 むき 込み処理を行った順にシーケンス番号を付与し、これと 前記データが母き込まれた論理ポリューム(第1の記憶 資源)上の位置(格納位置)を特定する即き込み位置情 報とを対応づけて (所定のテーブルに) 配修する (512 プ方式によるデータ差分管理の詳細について説明する。 2)。なお、母き込み位置情報は、例えば、セクタ番 <u>-</u> 9

ともに記憶サブシステム2に送倒する (SI23). ここで 送信した後に行われ、また、このコマンドには必要に応 【0154】 つぎに、配憶サブシステム1は、前配借込 対象データを、これに付与された前記シーケンス番号と このような記憶サブシステム間で行われる、データとジ **ーケンス番号の送信は、例えば、データ送信コマンドを** じて、前述したデータの掛き込み位配情報が付帯され

号、トラック番号等を用いて記述される。

記むき込み処理が完了すると、その完了通知を記憶サブ [0155] 配像サブシステム2は、配像サブシステム 1から送られてくる前配告き込み対象データとシーケン ス番号とを受信して、これを自身の論理ボリューム(第 2の記憶資獻)に你き込む。記憶サブシステム2は、前 システム1に送信する。 20

52

[0156] 配像サブシステム2は、配像サブシステム ⊠ 3に対し、適宜なタイミングで前記事き込み対象データ 16 (b) では、時間差を表現するため、記憶サブシス テム1から配位サブシステム2に送信されるデータのシ ーケンス番号と、配像サプシステム2から記憶サブシス テム3に送信されるデータのシーケンス番号とを変えて と前記シーケンス番号とを送信する (5124) (なお、 扱語している)

3

タと前記シーケンス番号とを受信すると、前記むき込み 配位サブシステム1に送信する (S125) 。 記位サブシス テム1は、配位サブシステム3から送られてくるシーケ 【0157】つぎに、記憶サブシステム3は、前紀デー 対象データに対応して発行した前記シーケンス部号を、 ンス番号を受信する。 32

ーケンス番号と、自身が配憶しているシーケンス番号と [0158] ここで記憶サブシステム1は、受信したシ ち、쒚分データを把握することができる。なお、前配の これに対応する母き込み位置情報との対応づけ(テープ ル)を対照することで、配位サブシステム3の論理ポリ 対照は、例えば、記憶サブシステム3から受領した事き ューム (第3の記憶資額) に未反映のデータ、すなわ 40

込み完了位置までのシーケンス番号と書き込み位置情報 とをテーブルから削除することにより行われる (512 45

[0152] つぎに、図16 (b) とともにマルチホッ 50 [0159] 以上のようにしてマルチホップ方式におけ

- 18 -

る通常運用が行われる。

[0160] つぎに、災害等により記憶サブシステム2 が停止した場合の回復処理について説明する。

[0161]図17 (a)に示すように、記憶サブシス テム1は、例えば、ハートビートメッセージの監視など の障害検出機能により、配憶サブシステム2の稼働状態 が記憶サブシステム2の障害発生を検知した場合に、記 **憶サブシステム1と記憶サブシステム3の間を、差分デ** をリアルタイムに監視している。以下では、ハートピー トメッセージが途切れるなどして、記憶サブシステム1 一クのみを複写することによってその内容を一致させ、 その後記憶サブシステム1と記憶サブシステム3の間 を、非同期転送での臨時運用へ移行させる処理につい 図17 (b) とともに説明する。

2 [0162] 記憶サブシステム1は、記憶サブシステム ケンス番号と前記書き込み位置情報との対応づけに基づ リ6上に、自身の論理ポリューム(第1の記憶資源)の 所定プロック単位のデータ格納位置に対応づけたビット マップを生成し、自身が記憶している記憶サブシステム 2の障害発生を検知した場合 (S131) 、まず、制御メモ 3 において未反映の前記差分データについての前記シー いて、デーク更新のあった前記ピットマップに対応する 位置のピットをオンにする (\$132)。

ームの、前記ピットマップ上のオンになっている位置に 格納されている差分データを、記憶サブシステム1から 33)。そして、この複写完了後、記憶サブシステム1か ら非同期転送により差分データが複写される形態で、臨 [0163] つぎに、記憶サプシステム1の論理ポリュ 記憶サプシステム3の対応する格納位置に複写する (SI 時運用が開始される (S134)。

【0164】ここでこの臨時運用への切り替えに際して

は、記憶サブシステム2に障害が発生した場合でも、記 億サブシステム1のデータを記憶サブシステム3に全部 い。このため、例えば、記憶サブシステム1と記憶サブ システム3との間の通信回線のデータ伝送量が充分でな い場合でも、各記憶サブシステムにおける論理ポリュー ムに記憶されているデータを容易に同期させることがで **複写する必要がなく、差分データのみを複写すればよ**

[0165] つぎに、配憶サブシステム2が復旧し、臨 時運用から通常運用に切り替える際の一連の処理につい

ータを記憶サブシステム2の論理ポリューム (第2の記 ブシステム2にデータを複写する同期転送での運用を開 にデータ費き込みを行った場合、書き込んだデータとシ 【0166】まず、記憶サブシステム1は、自身の論理 飽資額) に複写した後、記憶サブシステム 1 から記憶サ 始する。すなわち、記憶サブシステム1は、ホストから ポリューム(第1の記憶資源)に記憶している全てのデ の指示により自身の論理ポリューム(第1の記憶資源)

-ケンス番号とを記憶サブシステム2に送信する。

[0167] 記憶サブシステム2は、記憶サブシステム 1から送られてくる前記書き込んだデータとシーケンス 番号とを受信して、これを自身の論理ボリューム (第2 の記憶資源)に書き込む。記憶サプシステム2は、前記 書き込み処理が完了すると、自身の論理ポリューム (第 2の記憶資源)に対するデータ費き込みが行われた位置 を特定する書き込み位置情報と、データの書き込み順に ブルに) 記憶する。この段階のデータ転送状態を図18 付与されるシーケンス番号とを対応づけて (所定のテ

記対応づけにおける前記シーケンス番号を記憶サブシス 【0168】 つぎに、記憶サブシステム3は、記憶サブ システム 1 から送られてくる前配データと前配シーケン ス番号とを受信して、前記データを自身の論理ポリュー ム (第3の記憶資源) に記憶するとともに (図18) 前 テム2に送信する(図示せず)。

[0169] 記憶サブシステム2は、記憶サブシステム と、自身が記憶しているシーケンス番号と、これに対応 する暮き込み位置情報とを対照することで、記憶サブシ 3から送られてくるシーケンス番号を受信する。 ここで 記憶サプシステム2は、前記受信したシーケンス番号 ステム3の論理ポリュームに未反映のデータ、すなわ ち、差分データを把握することができる。

用を停止する。この停止後、記憶サブシステム2は、自 身の制御メモリ上に、自身の論理ボリューム(第2の記 [0170] つぎに、臨時運用において記憶サブシステ ム1から記憶サブシステム3に複写する非同期転送の運 **憶資源)の所定プロック単位のデータ格納位置に対応づ** けたビットマップを生成し、自身が記憶している記憶サ のシーケンス番号と書き込み位置情報との対応づけに基 づいて、データ更新のあった前記ピットマップの該当位 プシステム 3 において未反映の前記差分データについて 置のビットをオンにする。 25 30

リューム(第3の記憶資源)において未反映となってい [0171] つぎに、記憶サブシステム2は、前記ビッ トマップにより把握した、記憶サブシステム3の論理ポ る差分のデータとその書き込み位置情報とを記憶サブシ ステム3に送信する。 35

を、自身の論理ポリューム(第3の記憶資源)の、前記 **蹬に記憶する。これにより、記憶サブシステム2の論理** テム3の論理ポリューム(第3の記憶資源)の内容との [0172] 記憶サブシステム3は、前記差分データと 書き込み位置情報により指定される該当データの格納位 ボリューム(第2の記憶資源)の内容と、記憶サブシス 同期が取れることになる。以上の処理終了後、記憶サブ システム2と記憶サプシステム3との間の非同期転送に よる運用が開始され、図19に示す通常状態での運用が 前記費き込み位置情報とを受信して、前記差分データ 再開する。 **\$**0 45 20

[0173]以上のようにして臨時週用から通常週用へ

の切り替えが完了する。

【0174】 <マルチコピー方式>図20に示す広域デ ータストレージシステムは、サイト1に設置された記憶 サブシステム1と、サイト2に設置された記憶サブシス テム2と、サイト3に設置された記憶サブシステム3と を備える。記憶サプシステム2にはこの記憶サプシステ お、記憶サプシステム1と記憶サブシステム3との間も ム2を記憶手段として利用するホストが接続する。な 通信手段により接続される。

一夕を複写する同期転送で運用されている。また、記憶 ステム2から記憶サプシステム3にデータを複写する非 テム間の通信を同期転送とするか、非同期転送とするか は前記の形態に限られず、任意に設定される。また、同 [0175] 記憶サブシステム1と記憶サブシステム2 とは、記憶サブシステム2から記憶サブシステム1にデ サブシステム2と記憶サブシステム3とは、記憶サブシ る。なお、マルチコピー方式において、各記憶サブシス タ差分管理方式について説明する。 記憶サブシステム2 は、ホストから替き込み対象データとその書き込み要求 (Write 1/0) とを受信すると (S161) 、 番き込み対象 して同時に、前記書き込まれたデータと前記付与したシ **目した後に行われ、また、このコマンドには、必要に応** 書き込み位置情報は、例えば、セクタ番号、トラック番 同期転送で運用されている。以下、このような形態のリ データを自身の論理ポリューム(第2の記憶資源)に曹 き込む。また、記憶サブシステム2は、僣き込まれたデ **一夕と、書き込み処理を行った順に付与したシーケンス** 番号とを、記憶サブシステム1に送信する(8162)。そ ケンス番号の送信は、例えば、データ送信コマンドを送 位置(格納位置)を特定する書き込み位置情報とを対応 [0176] つぎに、図20とともにこの実施例のデー ーケンス番号とを、記憶サブシステム3に送信する (SI 【0177】 つぎに、記憶サプシステム1は、記憶サブ シーケンス番号とを受信して、前記書き込み対象データ が書き込まれた論理ポリューム (第1の記憶資源) 上の 64)。 なお、前述のマルチホップ方式の場合と同様に、 このような記憶サブシステム間で行われるデータとシー システム2から送られてくる前記書き込み対象データと ひ。その際、前記シーケンス番号と、これと前記データ じて前述したデータの書き込み位置情報が付帯される。 づけて (所定のテーブルに) 記憶する (S163)。 なお、 モートコピー制御方法を「マルチコピー方式」と称す を自身の論理ポリューム(第1の記憶資源)に書き込 彻転送や非同期転送以外の転送方式であってもよい。 号等を用いて記述される。

ンステム2から送られてくる前記書き込み対象データと 【0178】 つぎに、記憶サブシステム3は、記憶サブ シーケンス番号とを受信して、前記書き込み対象データ を自身の論理ポリューム(第3の記憶資源)に書き込

ム1に送信する (8165) 。 記憶サブシステム1は、記憶 サブシステム1に対し、前記書込み対象データとこれと 対になっていた前記シーケンス番号とを記憶サブシステ サプシステム3から送られてくるシーケンス番号を受信 む。番き込みが完了すると、記憶サブシステム3は記憶

O

0

.LΩ ~

2

-

特開2003

リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

【0.179】ここで記憶サブシステム1は、前記受信し たシーケンス番号と、自身が配像しているシーケンス番 号とこれに対応する告き込み位置情報との対応づけを対 照することで、記憶サプシステム3の論理ポリューム (第3の記憶資源)に未反映のデータ、すなわち、差分 例えば、記憶サブシステム3から受領した書き込み完了 データを把握することができる。なお、前記の対照は、 位置までのシーケンス番号と書き込み位置情報とをテ ブルから削除することで行われる (\$166)。 2

[0180] 以上のようにしてマルチコピー方式におけ る通常運用が行われる。 [0181] つぎに、災害等により記憶サブシステム2 が停止した場合の回復処理について説明する。

[0182] 図21 (a) に示すように、記憶サブシス テム1は、例えば、ハートピートメッセージの監視など の障害検出機能により、記憶サブシステム2の稼働状態 をリアルタイムに監視している。以下では、ハートビー トメッセージが途切れるなどして、記憶サブシステム1 20

が記憶サブシステム2の障害発生を検知した場合に、記 システム1と記憶サブシステム3の間を、差分データの 記憶サブシステム1と記憶サブシステム3の間を、非同 みを複写することによってその内容を一致させ、その後 **憶サブシステム2に接続するホストに代えて、記憶サブ** 切転送での臨時運用へ移行させる処理について、図21 52 30

[0183] 記憶サブシステム1は記憶サブシステム3 **タの操作により、記憶サブシステム2に接続していたホ** ストの業務の運用が、記憶サブシステム1に接続する副 の障害発生を検知した場合(S171)、例えば、オペレ ホストに引き継がれる。

(b) とともに説明する。

リ6上に、自身の論理ポリューム(第1の記憶資源)の マップを生成し、自身が記憶している記憶サブシステム [0184] つぎに、記憶サブシステム1は、制御メモ 所定ブロック単位のデータ格納位個に対応づけたピット 3 において未反映の前記差分データについてのシーケン デーク更新のあった前記ピットマップの該当位儲のピッ ス番号とデータ更新位置情報との対応づけに基づいて、 トをオンにする (SI72)。

【0185】 つぎに、記憶サブシステム1の論理ポリュ **一ムの、前記ピットマップ上のオンになっている位置に** 対応する位置に格納されている差分データを、記憶サブ 3)。そして、複写完了後、記憶サブシステム1から同 システム1から記憶サブシステム3に複写する (SI7 45

20

始される (SI74)。 [0186] ここでこの臨時運用への切り替えに際して

は、配徳サブシステム2に応導が発生した場合でも、記憶サブシステム1のデータを記憶サブシステム3に全部 複写する必要がなく、逆分データのみを複写すればよい。このため、倒えば記憶サブンステム1と記憶サブシステム3との間の通信回線のデータ伝送肌が先分でない場合でも、各記憶サブシステムにおける論理ポリューム場合でも、各記憶サブシステムにおける論理ポリュームに記憶されているデータを簡単に回溯させることができ

【0187】 つぎに、記憶サブシステム2が復旧し、臨時週用から通常週用に切り替える際の一連の処理につい

て説明する。

【0188】まず、記憶サブシステム1は、自身の論理 ポリューム (第1の記憶受演) に記憶している全てのデ ータを記憶サブシステム2の論理ポリューム (第2の記 億資額) に複写した後、記憶サブシステム1から記憶サ ブシステム2にデータを複写する同場転送での選用を明 始する。なお、このとき記憶サブシステム1と記憶サブ システム3周での非同期転送を継続して行われる。 [0189] 配億サブシステム1は、ホストから窃き込まれたデータと、印き込み処理を行った順に付与した・一ケンス番号とを、配億サブシステム2に送信する。そして同時に、前起むき込まれたデータと前記付与したシーケンス番号とを、配億サブシステム3にも強の論理ポリューム(第2の配億資源)に対するデータ母き込みが行われた位置を特定する母き込み位置情報と、データの母き込み項に付きされるシーケンス番号との対応づけを記憶する(位置情報管理デーブル作成)。この段階での適用状態を図22に示す。

必要でなくなる。

[0191]記憶サブシステム3は、配億サブシステム1から送られてくる前期データと前記シーケンス毎号とを受信して、前記データを自身の論型ボリューム(第3の記憶改選)に記憶するとともに前記対応づけにおける前記シーケンス希号を記憶サブシステム2は、記憶サブシステム3から送られてくるシーケンス希号を受信する。この記憶サブシステム2は、前記受信ルシーケンス番号を受信する。この記憶サブシステム2は、前記受信ルシーケンス番号と、自身が記憶している前記対応づけを対照することで、記憶サブシステム3の論型ポリュームに末反映のディーが、記憶サブシステム3の論型ポリュームに末反映のディーをからに

て、配切ケノアイナム3の編集がリュームに未攻映のアータ、すなわち、差分データを把握することができる。 【0193】つぎに、臨時週用において記憶サブシステム1から記憶サブシステム3に複写する非同期転送の週 用を停止する。この停止後、記憶サブシステム2は、自身の衝撃メリューム (類2の記憶投資)の所定プロック単位のデータ格特位配に対応づけたビットマップを生成し、自身が記憶している記憶サブシステム3において未反映の前記差分データについてのシーケンス番号とむき込み位置情報との対応づけに表

ゴいて、データ更新のあった前配ビットマップの核当位 図のビットをオンにする。

ロン・ノーロインによって。 「0194] つぎに、配位サブシステム2は、前記ピットマップにより担接した、配位サブシステム3の論理ボ 05 リューム (第3の配像資源) において未反映となってい る差分のデータとそのむき込み位置情報とを記憶サブシ ステム3に送信する。 (0195) 記憶サブシステム3は、前記差分データと前記むき込み位配格報とを受信して、前記差分データを前記むき込み位配格報に基づいて自身の論理ポリューム(第3の記憶資額)に記憶する。これにより、記憶サブシステム2の論理ポリューム(第2の記憶資額)の内容と、配徳サブシステム3の論理ポリューム(第3の記憶資数)の内容と、配徳サブシステム3の論理ポリューム(第3の記憶資数)の内容と、記憶サブシステム2から記憶サブシステム3への非同期転送サブシステム2から記憶サブシステム3への非同期転送が開始される。この段階での適用状態を図23に示

[0196] ここで記憶サブシステム1に接続するホストの記憶サブシステム1へのデータ母き込み処理が完了しており、記憶サブシステム1と記憶サブシステム2の同期が取れている時に、記憶サブシステム1から記憶サブシステム2に対して行っていたデータの複写を、記憶サブシステム2から記憶サブシステム1に対して行うように切り替える。すなわち、同期が取れている状態で切りを表を行うことで、差分データを複写する等の作業が、り替えを行うことで、差分データを複写する等の作業が

[0197] つぎに、記憶サブシステム1に接続するホストにより運用されている業務を、記憶サブシステム2に接続するホストに引き継ぐ。そして、記憶サブシステ30 ム2から記憶サブシステム3にデータを複写する同期転送による運用を開始することで、図24に示す通常状態送による運用を開始することで、図24に示す通常状態

での運用が再開することになる。 【0198】以上のようにして臨時運用から通常運用への切り替えが完了する。

(0199) <他の降事役旧方式>つぎに、降害役旧方式のバリエーションについて説明する。

[0200] 図25に示すマルチホップ方式において、配能サプシステム1がダウンした場合(a) には、記憶サプシステム2に倒ホストを接続し、この倒ホストによりの記憶サプシステム1に接続するホストの業務を引き継く。なお、配憶サプシステム2と記憶サプシステム3の間では、非同期低送での運用が行われている(b)。

まず、記憶サブシステム2の全データを記憶サブシステム1に 45 ム1に複写し、別ホストの業務を記憶サブシステム1に 接続するホストに引き継ぐ。そして、前記の要領で、記 億サブシステム1と記憶サブシステム2との間のデータ 低送方向を逆向きにすることにより、道体週用を再関す

【0201】配億サプシステム1が復旧した場合には、

[0202] 図26に示すマルチホップ方式において、

5 (c)

2

記憶サブシステム3に砕沓が発生した場合 (a) には、記憶サブシステム3の仮旧後、記憶サブシステム2から配像サブシステム3に全データを複写して記憶サブシステム3のデータを記憶サブシステム2と同期させ、記憶サブシステム1から記憶サブシステム2にデータを複写 0する同期転送および記憶サブシステム2から記憶サブシステム3にデータを複写する非同期転送による過常週用を再開する(b)。

[0203] 図27に示すマルチコピー方式において、記憶サブシステム1に降車が発生した場合 (a) には、記憶サブシステム1の位旧後、記憶サブシステム2から記憶サブシステム1に全データを複写して記憶サブシステム1のデータを記憶サブシステム2から記憶サブシステム2から記憶サブシステム2から記憶サブシステム2から記憶サブシステム2から記憶サブシステム2から記憶サブシステム3にデータを複写する非同期転送による適常運用を非エルス。

[0204] 図28に示すマルチコピー方式において、記憶サブシステム3に辞すが発生した場合には、記憶サブシステム3の位旧後、記憶サブシステム2から記憶サブシステム3に全データを複写して記憶サブシステム3のデータを記憶サブシステム2から記憶サブシステム1にデータを複写する同期転送および記憶サブシステム2から記憶サブシステム3にデータを複写する同期に送および記憶サブシステム2から記憶サブシステム3にデータを複写する目別に近かるを複写する目別に近かる記憶サブシステム3にデータを複写する非同期転送による通常通用を再開

[0205] <複写元・複写先、書き込み位置储頼の官理>記憶サブメテム間でデータを転送する場合、データの転送元や転送先の設定や、その転送が同期・非同期いずれの方式で行われるかといった設定は、オペレータが各配億サブシステムを操作して設定する場合(なお、この場合には、倒えば、ある記憶サブシステムが向音を起して模えなくなった場合に、どの記憶サブシステムが、なのデータの転送元になり、どの記憶サブシステムが、なのデータの転送元になり、足の記憶サブシステムが、方のデータの転送元になり、記憶サブシステムが、が配売・フステムがはの配送先になるのかということを、システムの構成時に予め登録しておく)、記憶サブシステムに付番するシステムが自動的に行ようにしている場合など、システムの構成に応じて様々な形態で行われる。

[0206]また、シーケンス部号とむき込み位置情報の対応づけの管理は、例えば、オペレータが、転送元を転送法を記憶サブシステムに登録する操作を開始する契機で行う。

[0207] <記憶サブシステムの選択方式>図29に示す広域データストレージシステムは、記憶サブシステム1と1とこれに接続するホスト1h、記憶サブシステム1からデータが非同期伝送される記憶サブシステム2および記憶サブシステム3を縮えている。ホスト1nもしくは記憶サブシステム1に降音が発生した場合、出遊に記憶サブシステム2もしくは記憶サブシステム3のどちらか一方を主たる記憶サブシステムとして選択し、また、

信頼性・保全性確保のため、これら2つの記憶サブシステム2および3においてデータを2 近化管理する。以下、ホスト1hもしくは記憶サブシステム1に降百が発下、ホスト1hもしくは記憶サブシステム1に降百が発

生した場合に行われる処理について説明する。 【0208】配位サブシステム2は、例えば、配位サブシステム1から送信されてくるデータの有無や、配位サブシステム1からあらかじめ設定された時間等に送られてくるハートピートメッセージの監視により、ホスト1トや記憶サブシステムに降音が発生したことを検知す (0209) 阿存を役知した場合、配位サブシステム2 は、迅速に主たる配位サブシステムを決定し、耐ホスト 2もしくは倒ホスト3による臨時迎用に切り移える。主 たる記憶サブシステムの選択はつぎのようにして行われ 15 る。まず、障碍を役知した配億サブシステム2は、配億 サブシステム3に、前述したシーケンス帝号のうち吸筋 のシーケンス帝母の送信を契求するメッセージを送信す る。記憶サブシステム3は、前記メッセージを受信す と、自分が記憶している最第のシーケンス帝号を記憶サ 20 ブシステム2に送信する。

(0210)配値サブシステム2は、記憶サブシステム3から送られてきたシーケンス都号と、自分が記憶している最新のシーケンス都号とを比較して、より最新のシーケンス都号をを比較して、より最新のシーケンス都号を受信している配値サブシステムを、並たステムの職別子を選出候補として記憶する。記憶サブシステム3は、送信されてきた前記職別子を受信し、これによりどの記憶サブシステムがまたるサブシステムとして設出されたのかを認知する。

10211] なお、以上の遠出処理において、記憶サブ(0211]なお、以上の遠出処理において、記憶サブシステム間の適信方式の性質などの諸事情により、記憶・サブシステム2もしくは配憶サブシステム3が記憶しているシーケンス番号に抜けが存在することがある。そこい、このような場合には、連続しているシーケンス番号

のうちで母新のものを、前記の比較に用いる。 [0212] 主たる配像サプシステムが強川されると、 つぎに、記像サプシステム2と配像サプシステム3とに よりデータの二肌化管理を行うため、両名が配像してい 40 るデータの内容を一致させる。これは、配位サブシステ ム間で全データの複写や整分データの複写により行われ る。配像サブシステム間でデータが一致すると、主たる

配像サブシステムとして選出された配位サブシステムは、自分に接続している例ホストに、自分がまたる配位45 サブシステムとなる旨を送信する。例ホストはこれを受信して代行週用を開始する。また、配位サブシステム3との間で、同項転送もしくは非同と記憶サブシステム3との間で、同項転送もしくは非同

期転送によるデータの二瓜化管型が開始される。 【0213】なお、以上の裁明では、記憶サブシステム 2が記憶サブシステム3から段新のシーケンス部号を収

20

のデータを管理する単位であるプロックの概念を示す図

得して、主たる記憶サプシステムを選出するようにして いるが、この処理は記憶サブシステム3が行ってもよ

ステム3の3台構成の記憶サブシステムにおいて、記憶 サブシステム1の障害発生時に代行して運用される他の 記憶サブシステムを選出する仕組みを一例として説明し **構成される広域データストレージシステムにも適用する** [0214]また、記憶サブシステム1乃至記憶サブシ たが、前述の仕組みは、4台以上の記憶サブシステムで

[0215] <キャッシュメモリ上のデータの管理>ホ ストが接続する一次の記憶サブシステムに、この一次の 一次の記憶サブシステムのキャッシュメモリ上のデータ 記憶サブシステムのデータのリモートコピー先である1 以上の二次の記憶サブシステムが接続する系における、 の管理に関する実施例について説明する。

システムの記憶資源にデータを書き込んだ後は、そのデ ステムに送信するまではキャッシュメモリ上に残してお 転送が終了するまで、データをキャッシュメモリ上に残 【0216】前紀の系において、一次の記憶サブシステ 一)する必要の無いデータについては、一次の配像サブ 一夕を当該配像サプシステムのキャッシュメモリ上から 消去されてもよいが、二次の記憶サブシステムに複写す る場合には、少なくともそのデータを二次の記憶サブシ く必要がある。また、転送先となる二次の記憶サブシス テムが複数存在する場合には、通信手段の違いや運用上 の差異などにより、通常、二次の記憶サブシステムにつ いての転送が同時に行われるわけではないので、このよ うな場合には、全ての二次の記憶サブシステムに対する ムから二次の記憶サプシステムに複写(リモートコピ しておく仕組みが必要である。

ムについての転送が完了しているかどうかを管理するよ 転送が完了しているかどうかを示すテーブルを、一次の て、キャッシュ上に置かれているデータについて、一次 ているデータについて、二次の各記憶サブシステムへの の記憶サブシステムに接続する二次の各記僚サブシステ ~;#n)ごとに、それぞれの記憶プロックに格納され うにする。具体的には、例えば、図30に示すように、 [0217] そこで、一次の記憶サブシステムにおい キャッシュメモリ上に区画された記憶プロック (#1, 記憶サブシステムにおいて管理するようにする。

プシステムについてのピットは、転送完了後に「0」と トされる。ある記憶プロックの「1」がセットされてい、 るビットのうち、データの転送が完了した二次の記憶サ タが一次の記憶サブシステムに書き込まれた時に、デー 次の記憶サブシステムに対応するビットに「1」がセッ タが掛き込まれた配像プロックの転送先となっているニ 「0」は転送が完了していることを示し、ビット「1」 は転送が完了していないことを示す。ホストからのデー [0218] なお、このテーブルにおいて、ピット

れているデータについては、キャッシュメモリ上から消 【0219】そして、全ての二次の記憶サブシステムに ついてのビットが「0」となった記憶プロックに格納さ 去してもよいということになる。

上のサイトを有する広域データストレージシステムにお も、巨視的に見て常時、データの順序性を保証した論理 [発明の効果] 図1、図9及び図10で示した、3つ以 いて、いずれかのサイトに、いつ災害・障害が発生して

ポリュームを残すことができる。

【0221】直接データ転送に関与していなかった論理 ば、即時に、非同期リモートコピーのペアを生成でき、 2

ポリューム間、例えば、図7の記憶サブシステム1と記 節サプシステム3との間で、差分データのみコピーすれ 広域データストレージシステムの運用再開が、即時に、 可能となる効果がある。

モートコピーを実施するための冗長な論理ポリュームを 必要としないため、記憶サブシステムのメモリー資源の 【0222】本発明では、記憶サブシステムの内部にリ 使用効率が上がり、記憶サブシステムのコストパフォー マンスが向上する効果がある。 20

[図1] 本発明に係る広域データストレージシステムの 全体構成の一例を示した説明図である

【図2】 記憶サブシステムの一例を示した概念図であ

[図3] 図1の構成において、データコピー監視機能を 説明するための概念図である

[図4] 本発明を実現するための転送状態/ビットマッ プの一例を示した図である。 30

[図5] 一般的な同期転送によるコピーの制御の概略を

説明するための図である。

【図6】非同期リモートコピーの制御の概略を説明する ための図である。 [図7] 図9の全体構成において、データセンタ2に障 資・災害が発生した場合の復旧の様子を示した説明図で、 [図8] 図1の全体構成において、データセンタ1に障 害・災害が発生した場合の復旧の様子を示した説明図で 40

[図9] 本発明に係る広域データストレージシステムの 全体構成の別の一例を示した説明図である。

本発明に係る広域データストレージシステムの全体構成 【図10】 データセンタを4拠点以上設置した場合の、 の別の一例を示した説明図である。 45

データコピー監視 【図12】本発明の一実施例による非同期転送における [図11] 図9の全体構成において、 機能を説明するための概念図である。

データの整合性の管理方法を説明するための、記憶資源

【図13】本発明の一実施例による非同期転送における データの整合性の管理方法を説明するための、データ管 理情報の概念を示す図である。

【図15】本発明の一実施例による非同期転送における データの整合性の管理方法を説明するための、記憶サブ 【図14】本発明の一実施例による非同期転送における データの整合性の管理方法を説明するための、転送デー タのフォーマットの概念を示す図である。

システム2において管理されるデータ管理情報の概念を [図16] (a) はマルチホップ方式の広域データスト レージシステムの概念を示す図であり、(b)は(a)

に記載の記憶サブシステムにより行われる処理の流れを [図17] (a) はマルチホップ方式の広域データスト に記載の記憶サプシステムにより行われる処理の流れを フージシステムの概念を示す図であり、(b)は(a) 示す図である。

常運用に切り替え途中段階における記憶サブシステム間 [図18] マルチホップ方式において、臨時運用から通 のデータ転送状態を示す図である。

[図19] マルチホップ方式において、臨時運用から通 常運用への切り替え終了後の記憶サブシステム間のデー [図20] (a) はマルチコピー方式の広域データスト 夕転送状態を示す図である。

に記載の記憶サブシステムにより行われる処理の流れを レージシステムの概念を示す図であり、(b)は(a) 示す図である。

に記載の記憶サブシステムにより行われる処理の流れを [図21] (a) はマルチコピー方式の広域データスト レージシステムの概念を示す図であり、(b)は(a)

[図22] マルチコピー方式において、臨時運用から通 常運用に切り替え途中段階における記憶サブシステム凹 のデータ転送状態を示す図である。

6

0

Ŋ 7 ~

ī

特開2003

常運用に切り替え途中段階における記憶サブシステム間 [図23] マルチコピー方式において、 のデータ転送状態を示す図である 05

[図24] マルチコピー方式において、臨時巡用から通 常運用への切り替え終了後の記憶サブシステム間のデー マルチホップ方式におけ [図25] (a) ~ (c) は、 夕転送状態を示す図である。 10

る障害復旧方式の他のパリエーションを説明する図であ [図26] (a)、(b) は、マルチホップ方式におけ

る障害復旧方式の他のパリエーションを説明する図であ

[図27] (a)、(b) は、マルチコピー方式におけ る障害復旧方式の他のパリエーションを説明する図であ 'n. [図28] (a), (b) は、マルチコピー方式におけ る障害復旧方式の他のパリエーションを説明する図であ

[図29] 障害発生時において、本番業務を代行させる 記憶サプシステムの選択方式を説明する、広域データス

テムへのデータの転送状態を管理するテーブルを示す図 【図30】本発明の一実施例による、キャッシュメモリ 上のデータの管理方法における、二次の各記憶サプシス トレージシステムの競佘図である。

52

1 記憶サブシステムの制御装配 [符号の説明] 200

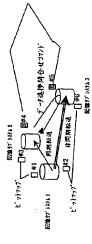
キャッシュメモリ

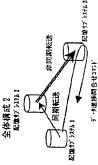
#1~#6 転送状態/ピットマップ。

[図3]

(6区)

デ-921. - 監視機能





20

509

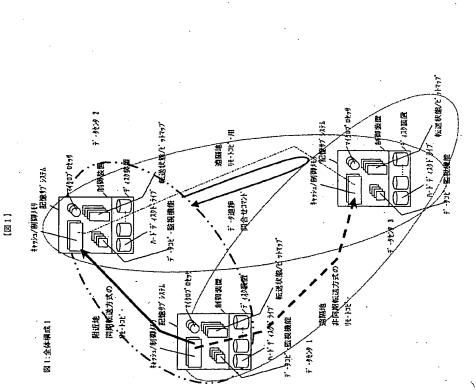
特開2003-122

[図2]

- 56 -

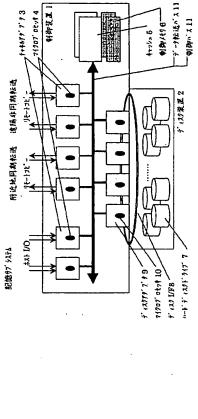
2003 07 04 17:30

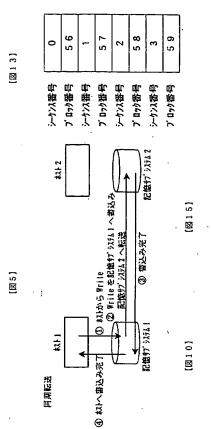


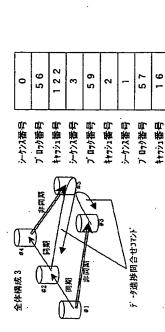


[図4]

航送状態/ピッパップ







[9図]

[図11]

- 21 -

1 からの # 1

[8 ⊠

[図7]

[图12]

ブロック番号56 ブロック番号51 プロック番号58 ブロック番号59

[図14]

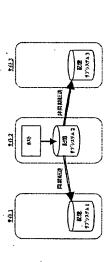
フロッケ番号 デ・4 フロッケ番号 データ フロッケ番号 データ シャンス番号 データ

特開2003-122509

リモートコピー制御方法、これを用いた記憶サブシステム、及び、これらを用いた広域デ...

[図17]

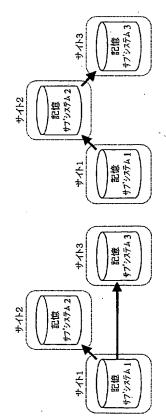
3



£-3 F=+3 (b) Wale 1/0 を気得(5161)

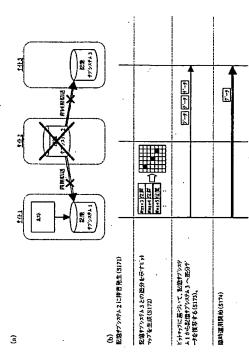
F=+5 2-9 92243 8=1+5 位 E 記憶サブシステム3から受信した 等込み完了位置までのゲーケンス等 サと音込み位置をサーブ ルから開除する(S166) 女信したゲーナをシーナンス番号 と共に記憶サブンスタム2 に設 信 (S162) 記憶すアンストム2はプークの音 込みを行う。 記憶すアンステムは交信したソ ーシンス部分と音込み位置は 得とを記憶する(S163) 記憶サブシステム21セデーナをシー シンス番 みと共に配給サブシステ ム3 に送信(5164) 記憶47°227431は7~9の各 込みを行う。また、ケーケンが 号を記憶47°22441に送信 する(9163)

[図19] [図18]



22l8 97'v.xf.k.3 F-3 F-9 F-9 1 記載 \$75.5.8.8.1 (b) 記憶サプシステム2 に降害発生(S131) ビントランド取るこれ、四世ブン メチューセル内のアンスティ3 〜如 タデーシを含みする(5133) **記値サプシステム3との益分をボ すピゥトャゥブを生成(S132) 商時運用開始(S134)** 3

[図21]



#218 #7": x#A 3

Î #=3+5 5:-9

受信したデークをシーケンス番号 と実に記憶サブ・ステム2 に送 (音(5162)

Write I/O を受信 (S161)

3

17'8'7'A'A Z

知識を記

#7'77#A 1

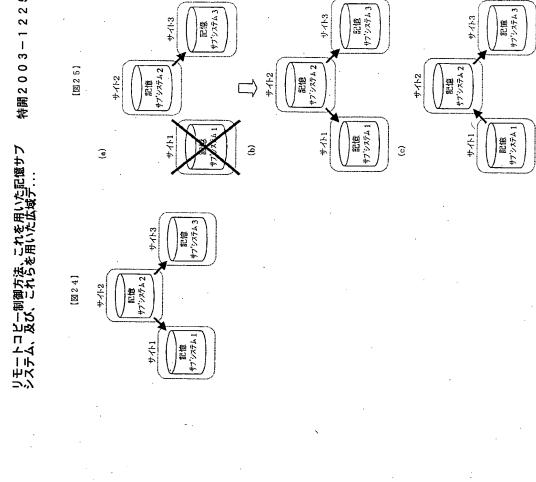
443

<u>\$41.2</u> *****

3

[図20]

509



#=8+5 F'-9

#=E+3

記憶サブシステム3はデークの雷 込みを行う。また、シーケス番 号を記憶サブシステム1に送信 する(S165)

配位サブンスチム2はデーナをシー ケンス母号と共に配修サブンステ ム3に送信(8164)

#=6+4位配 #==+5 位面 :

配替サアッポル3から受信した 参込み先了位置までのシーケン 本番号と尊込み位置をアープ ハから別称する(5166)

記憶 サブシステム3 記憶 サブシステム2 記憶 サブシステム 1 サイトコ 記憶 サブシステム3 #4F3 記憶 サブシステム 2 #4k2 記憶 サブシステム 1 44k1

[図23]

[図22]

- 34 -

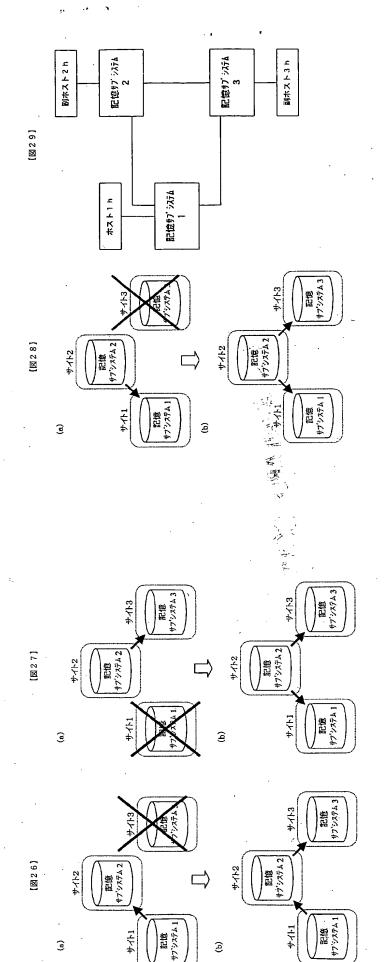
リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

特開2003-122509

リモートコピー制御方法、これを用いた記憶サブ システム、及び、これらを用いた広域デ...

တ

特開2003-12250



フロントページの統計

[図30]

(72)発明者 尾形 幹人 神奈川県小田原市中里322番地2号 株式 会社日立製作所RAIDシステム事業部内 (72)発明者 岡見 吉規 神奈川県小田原市中里322番地2号 株式 会社日立製作所RAIDシステム事業部内 (72)発明者 柏垣 錬一 神奈川県小田原市中里322番地2号 株式

会社日立製作所RAIDシステム事業部内 会社日立製作所RA1.Dシステム事業部内 神奈川県小田原市中里322沿地2号 株式 种奈川県小田原市中里322番地2号 株式 Fターム(参考) \$B065 BA01 CE01 EA35 5B082 DE03 GA04 (72)発明者 安部井 大 木瓶 斑 (72) 発明者 35

49

\$\$75,181 \$\$75,1\$2 \$\$75,1\$3 \$\$75,1\$4

1 33

BEST AVAILABLE COPY